Lipoylation and assembly of a 2-oxoacid dehydrogenase multienzyme complex from thermoplasma acidophilum

Posner, Mareike

Award date:
2009

Awarding institution:
University of Bath

Lipoylation and assembly of a 2-oxoacid dehydrogenase multienzyme complex from *Thermoplasma acidophilum*

Mareike G. Posner

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

April 2009

Acknowledgements

I would like to thank Professor Michael Danson, Dr. David Hough and Professor Richard Perham for giving me the opportunity to work on this project.

Thanks to Dr. Edward McManus, Dr. Lisa Jennings for help with the SPR, Dr. David Scott for carrying out the analytical ultracentrifugation experiments and Dr. Tim Woodman for help with NMR.

I am grateful to Dr. Stefan Bagby for carrying out protein NMR experiments and Dr. Jean van den Elsen.

My special thanks go to Dr. Abhishek Upadhyay for his wonderful support, advice, help and encouragement throughout my Ph.D.

No words will be enough to thank my family for supporting me in my decision to study abroad and being there when I needed them.

CONTENTS

CONTENTS

CONTENTS

Abstract

Energy generating processes like the citric acid cycle are a pivotal part of metabolism. Members of the 2-oxoacid dehydrogenase multienzyme complex (OADHC) superfamily feed into and act within the citric acid cycle. OADHCs are composed of three enzymes: 2-oxoacid decarboxylase (E1), dihydrolipoamide acyltransferase (E2) and dihydrolipoamide dehydrogenase (E3). Covalent attachment of lipoic acid (LA) to E2 is essential for overall OADHC activity. Although thought to be absent in Archaea, it has recently been found that *Thermoplasma acidophilum* has all the components for an active recombinant OADHC (Heath *et al.*, 2007). Recent studies have further suggested that *Tp. acidophilum* may have an enzyme to covalently attach LA to E2 (Sun *et al*., 2007; McManus *et al.*, 2006).

This work describes the cloning and recombinant expression of the *Thermoplasma* lipoate protein ligase (*Tp*. LplA), its C-terminal domain and a fusion protein composed of the above two proteins. Both proteins are required for lipoylation of E2 *in vitro*. For the first time, *in vivo* lipoylation of E2 in *Tp. acidophilum* cell cultures is also being reported.

The effect of lipoylation and temperature on the *Thermoplasma* OADHC assembly has also been studied. This study revealed the temperature dependence of the E2 core and the whole complex assembly. These findings are in line with the optimum growth temperature of *Tp. acidophilum*. Dynamic light scattering and analytical ultracentrifugation were used to determine the molecular mass of whole OADHC. The molecular mass was determined to be 5 MDa with an octahedral geometry of the E2 core.

The results of this work strengthen the assumption that these enzyme systems may have had or potentially have a role in the Archaea. This may hold further clues to the evolutionary relationship between the three kingdoms of life and the role of OADHCs/lipoylation in the Archaea. The temperature dependent assembly of the complex and thermostability of these proteins may also provide a model to study thermostability and protein-protein interactions at high temperatures.

List of abbreviations

ACP	Acyl carrier protein
AUC	Analytical ultracentrifugation
BCOADHC	Branched-chain 2-oxoacid dehydrogenase complex
BRE	TFB recognition element
CTD	C-terminal domain of *Tp.* LplA
DHLA	Dihydrolipoic acid
DLS	Dynamic light scattering
E1	2-oxoacid decarboxylase
E2	Dihydrolipoamide acyltransferase
E3	Dihydrolipoamide dehydrogenase
EDTA	Ethylenediaminetetraacetic acid
GOR	2-oxoglutarate ferredoxin oxidoreductase
His-CTD	His-tagged *Thermoplasma* CTD
His-LplA	His-tagged *Thermoplasma* LplA
ITC	Isothermal titration calorimetry
LA	Lipoic acid
LipA	Lipoic acid synthase
LipB	Octanoyl-acyl carrier protein : protein-*N*-octanoyltransferase
LplA	Lipoate protein ligase
M	Molecular mass
M_r	Relative molecular mass
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
ntCTD	non-tagged CTD
ntLplA	non-tagged LplA
OADHC	2-oxoacid dehydrogenase multienzyme complex
OAFOR	2-oxoacid ferredoxin oxidoreductase
OEPCR	Overlap extension polymerase chain reaction
OGDHC	2-oxoglutarate dehydrogenase complex
PDHC	Pyruvate dehydrogenase complex
POR	Pyruvate ferredoxin oxidoreductase
SD	Shine-Dalgarno
SE	Standard error
SPR	Surface plasmon resonance
TLC	Thin layer chromatography
***Tp.* LplA**	*Thermoplasma acidophilum* lipoate protein ligase
TPP	Thiamine pyrophosphate (also cocarboxylase)
VOR	2-oxoisovalerate ferredoxin oxidoreductase

CHAPTER 1
INTRODUCTION

1.1 The superfamily of 2-oxoacid dehydrogenase multienzyme complexes

2-oxoacid dehydrogenase multienzyme complexes (OADHCs) are large protein assemblies ranging in size between 10^6 to 10^7 Da. They catalyze key reactions in intermediary metabolism and have been shown to regulate the cellular redox state by directly influencing the NADH/NAD^+ ratio (Günther *et al.*, 2005; Bunik, 2003). Members of the OADHC superfamily are the pyruvate dehydrogenase complex (PDHC), which links glycolysis with the citric acid cycle by catalyzing the conversion of pyruvate to acetyl-CoA, the 2-oxoglutarate dehydrogenase complex (OGDHC) which catalyses the conversion of 2-oxoglutarate to succinyl-CoA within the citric acid cycle, and the branched-chain 2-oxoacid dehydrogenase complex (BCOADHC), which oxidatively decarboxylates the branched-chain 2-oxoacids produced by the transamination of valine, leucine and isoleucine (Fig. 1.1) (Perham and Packman 1989; Yeaman 1989).

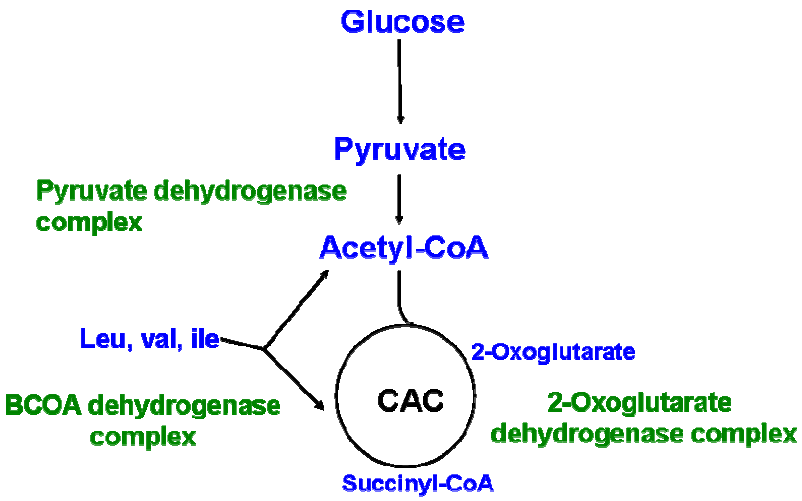


Fig. 1.1 Reactions catalyzed by the PDHC, OGDHC and BCOADHC. The main common feature is the oxidative decarboxylation of 2-oxoacids.

The PDHC, OGDHC and BCOADHC are made up of multiple copies of three component enzymes: a 2-oxoacid decarboxylase (E1), a dihydrolipoamide acyltransferase (E2) and a dihydrolipoamide dehydrogenase (E3). The substrate

specificity of each complex is determined by E1 and E2, whereas E3 is often the same gene product in different OADHCs (Danson *et al.*, 2004). The components E1, E2, and E3 assemble into a striking architecture dominated by the E2 core to which E1 and E3 bind non-covalently. The E2 core can either be composed of 24 or 60 E2 polypeptide chains arranged in an octahedral or icosahedral symmetry, respectively (Fig. 1.2).

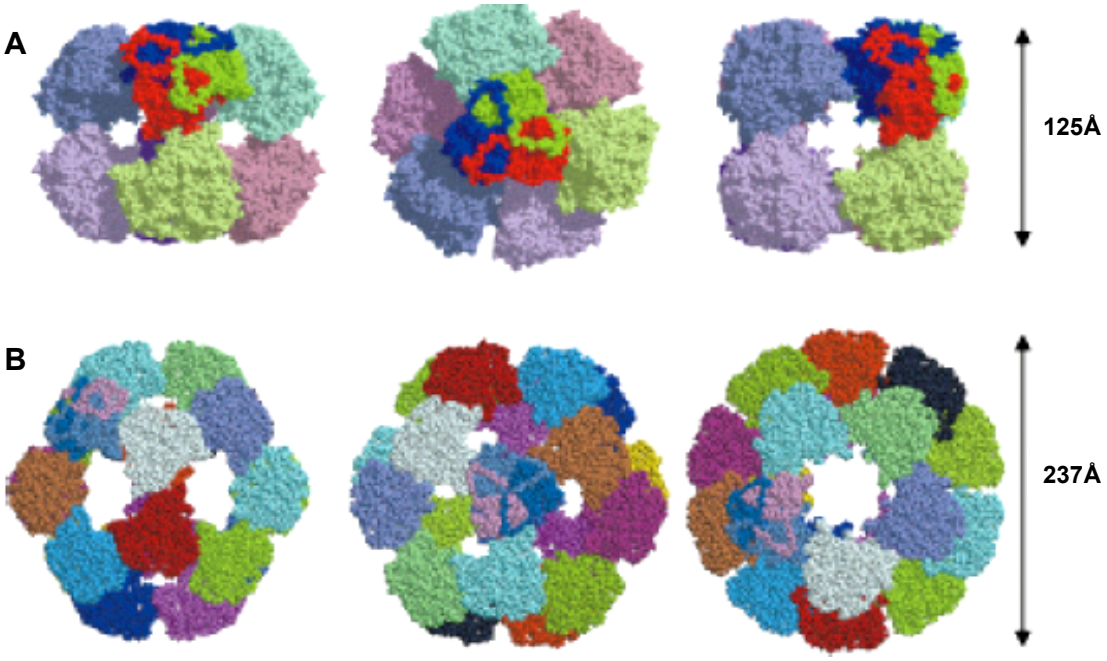


Fig. 1.2 Structures of the octahedral and icosahedral E2 inner cores (acyltransferase domains) of pyruvate dehydrogenase (PDH) complexes. **A**: E2 core of the *Azotobacter vinelandii* PDH complex; left to right, twofold, threefold, and fourfold views. **B**: E2 core of the *Bacillus stearothermophilus* PDH complex; left to right, twofold, threefold, and fivefold views. In both instances, the three different subunits of a trimer are in different colours, most readily seen on the threefold views (taken from Perham, 2000).

The quaternary structure of the E2 core is made possible by the compact subunit binding domain, which contains the binding sites for E1 and E3 as well as the catalytic site for transacylation. The E2 component of OADHCs contains one, two or three distinct lipoyl domains. Each of these domains has a distinct lysine residue to which the co-factor lipoic acid (LA) is attached covalently via an amide linkage between the carboxyl group of LA and the ε-amino group of the lysine residue forming a lipoamide moiety (Cronan *et al.*, 2005; Stepp *et al.*, 1981) (Fig. 1.3).

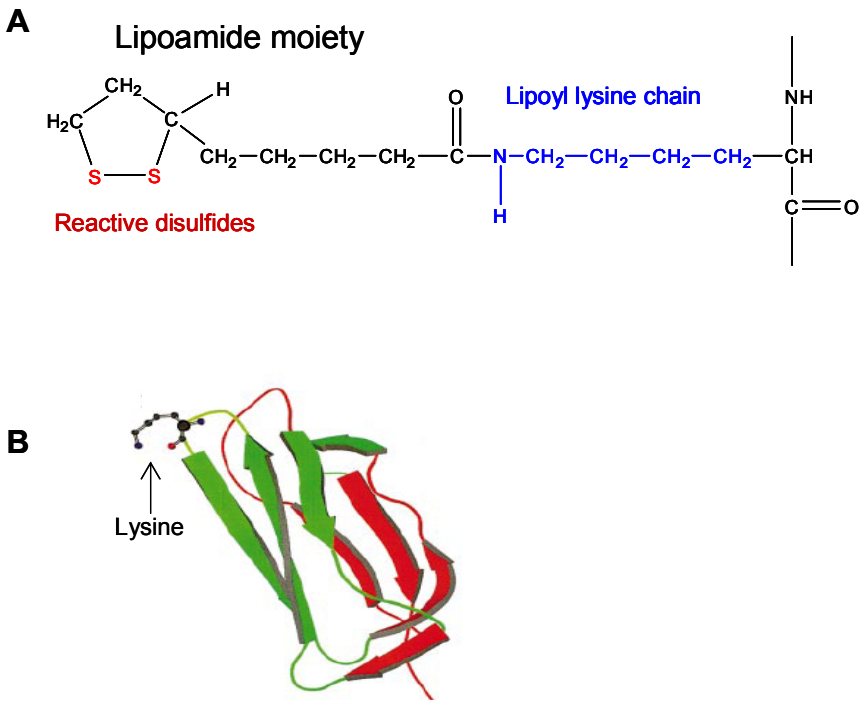


Fig. 1.3 Lipoic acid and the E2 lipoyl domain. **A**: Lipoamide moiety attached to the lipoyl lysine of the E2 lipoyl domain. **B**: The overall structure of the *E. coli* OGDHC E2 lipoyl domain; the lysine residue to which LA is attached is indicated (adapted from Perham, 2000).

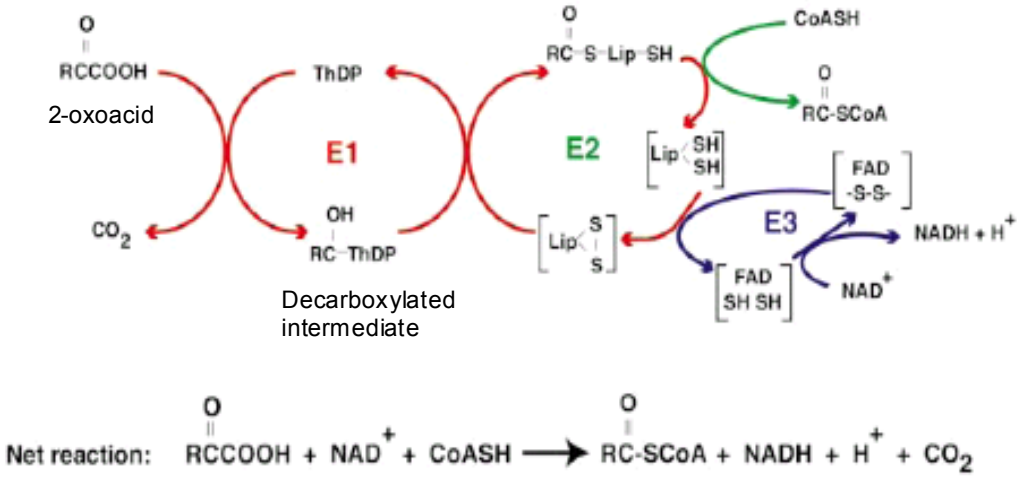


Fig. 1.4 General reaction scheme of the PDHC, OGDHC, and BCOADHC. This diagram illustrates the central role of the lipoyl moiety, able to take on the decarboxylated intermediate and passing it on to E2. The cofactors for E1, E2 and E3 are thiamine diphosphate (ThDP), lipoic acid and flavin adenine dinucleotide (FAD), respectively (adapted from Milne *et al.*, 2002).

The lipoamide moiety attached to the lipoyl domain is the key for the overall enzymatic activity of the OADHCs as it allows the specific transfer of reaction intermediates between the different enzyme active sites (Fig. 1.4) (Perham, 1991). The initial 2-oxoacid substrate is decarboxylated by E1. The consecutive steps are best visualized by considering the LA as a swinging arm that binds the decarboxylated intermediate via its reactive disulfide group. LA picks up the intermediate directly from the E1 active site. Then LA moves on and passes this intermediate to the active site of E2. In the final reaction step, LA interacts with the active site of E3 to allow regenerative oxidation of its reduced disulfide bond.

This mechanism copes well with the distance between different active sites within the multienzyme complex with high specificity while ensuring that the substrates and intermediates are protected during the whole reaction cycle. This has led to the well accepted concepts of active site coupling, substrate channelling and the "hot potato hypothesis" (Perham, 1991 and references therein; Miles *et al.*, 1988; Danson *et al.*, 1978a; Danson *et al.*, 1978b; Stanley *et al.*, 1981; Perham, 1975). Protein interactions that ensure specificity in the recognition of the lipoyl domain by E1 or E3 have been characterised for the *Bacillus stearothermophilus* PDHC (Fries *et al.*, 2007; Howard *et al.*, 2000).

1.2 Lipoic acid and 2-oxoacid dehydrogenase multienzyme complexes

Lipoic acid (LA) was first isolated and named by Lester Reed in 1951 (Reed *et al.*, 1951; Kresge *et al.*, 2006). LA is also known as 1,2-dithiolane-3-pentanoic acid and can exist in its reduced form as dihydrolipoic acid (DHLA) (Fig. 1.5). The ability of LA to cycle between its different redox states makes it an essential cofactor in enzymatic reactions involved in oxidative and single carbon metabolism (Cronan *et al.*, 2005; Zhao *et al.*, 2003). The naturally occurring R-isomer of LA is present in most prokaryotic and eukaryotic organisms as a physiological substrate whereas the S-isomer can only be synthesized chemically (Self *et al.*, 2000; Cronan *et al.*, 2005; Oehring and Bisswanger, 1992).

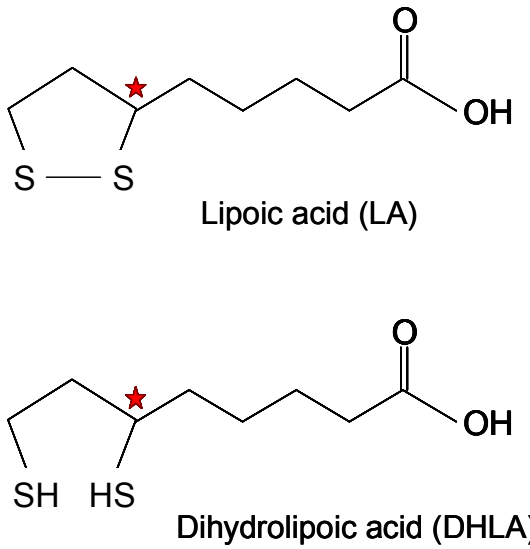


Fig. 1.5 Structure of lipoic acid and its reduced derivative dihydrolipoic acid (DHLA). The red star indicates the chiral centre of the two structures.

The model organism in which the synthesis of LA and lipoylation of proteins has been well studied is *Escherichia coli*. In *E. coli* protein bound LA can be derived from exogenous and/or endogenous stores using three enzymes: lipoate protein ligase (LplA), lipoic acid synthase (LipA) and octanoyl-acyl carrier protein (ACP):protein-*N*-octanoyltransferase (LipB) (Fig. 1.6). In the endogenous pathway LipB and LipA act consecutively to attach an octanoyl moiety to the enzyme's lipoyl domain and convert the octanoyl group into a lipoyl moiety respectively. In the exogenous pathway LplA first catalyzes the activation of exogenously supplied LA to lipoyl-AMP and then catalyzes the covalent attachment of LA to a protein lipoyl domain (Fig. 1.6). Despite the assumption that the three lipoylation enzymes are restricted to specific pathways, experiments with *E. coli* mutants showed that the enzymes are not mutually exclusive. LplA can bypass LipB activity by using octanoic acid as a substrate (Morris *et al.*, 1994; Zhao *et al.*, 2003; Cronan *et al.*, 2005) and LplA is able to use octanoyl-ACP as its substrate with low efficiency (Jordan and Cronan, 2003).

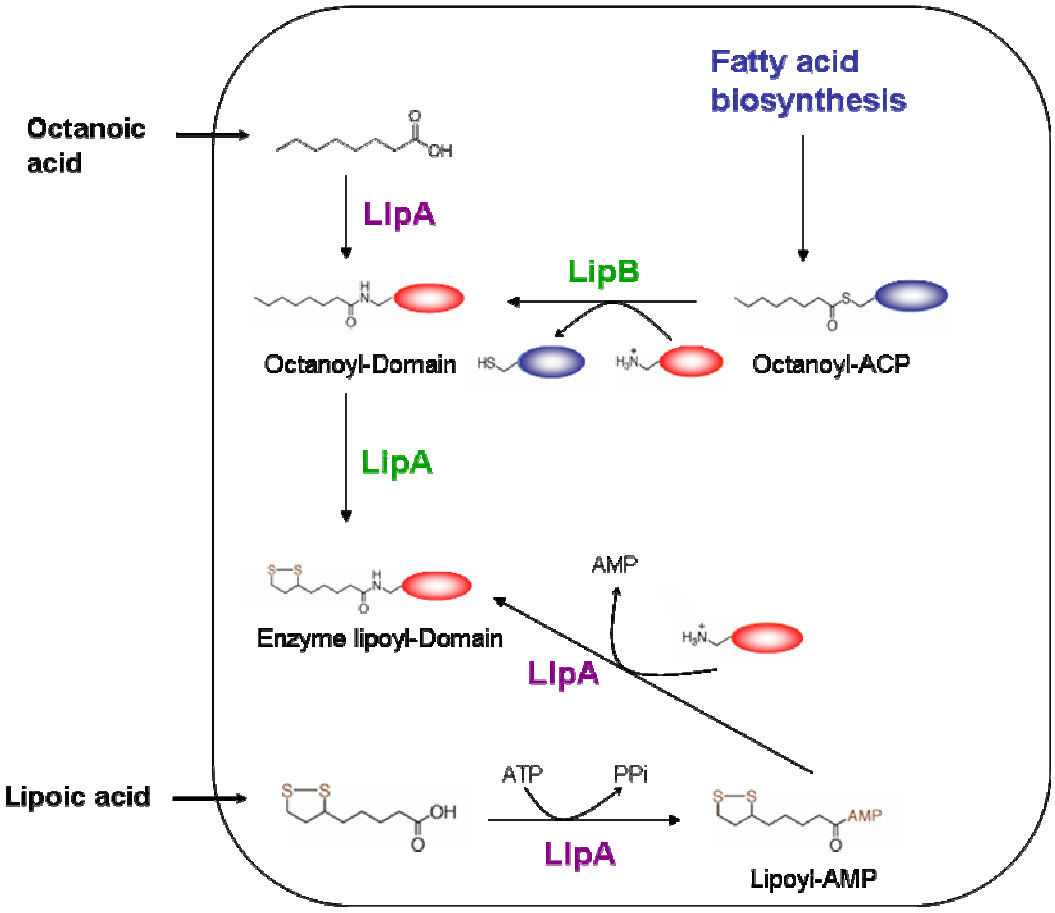


Fig. 1.6 Synthesis and attachment of LA in *E. coli*. The rectangle represents the *E. coli* cell. Octanoic acid and LA enter the cell by diffusion. Inside the cell, LA is attached to the lipoyl domain of the E2 of OADHCs (red ellipse). The enzymes LipB and LipA act in sequence and can use both endogenously or exogenously derived octanoic acid. In the endogenous pathway octanoic acid is presented on acyl carrier protein (blue ellipse) (adapted from Cronan *et al.*, 2005).

1.3 Archaeal 2-oxoacid dehydrogenase multienzyme complexes?

Archaea are classified as one of the three domains of life based on their rRNA sequence and cell membrane composition (Forterre *et al.*, 2002; Brown, 2003). Molecular analyses have shown that archaea share a common ancestral branch with eukarya although their metabolism more closely resembles that of bacteria (Pace, 2006; Forterre *et al.*, 2002; Chan *et al.*, 1997) (Fig. 1.7).

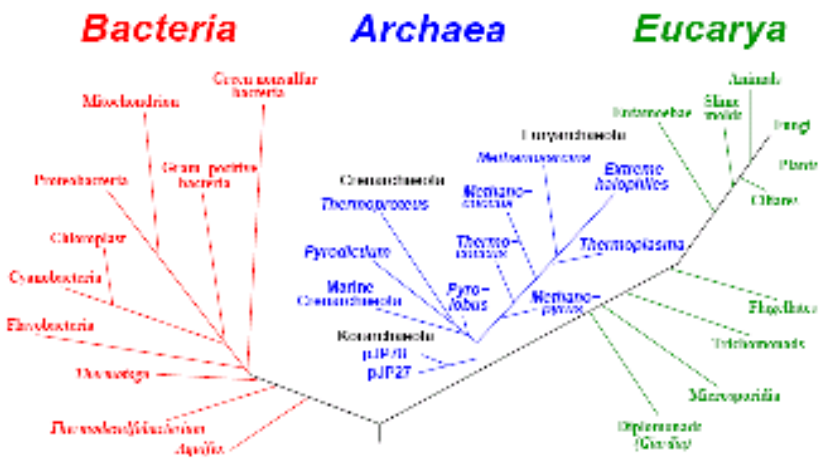


Fig. 1.7 Rooted universal phylogenetic tree showing the three domains based upon 16S (or 18S) rRNA sequences (from Woese *et al.*, 1990). The position of the root was determined by comparing paralogous gene sequences that diverged from each other before the three primary lineages emerged from their common ancestral condition (Iwabe *et al.*, 1989) (Jürgens, 2002).

Analyses of archaeal species have indicated that oxidative decarboxylation of pyruvate and 2-oxoglutarate under aerobic conditions is catalysed by 'anaerobic' 2-oxoacid ferredoxin oxidoreductases (OAFORs) instead of 'aerobic' OADHCs; therefore OADHCs are not needed in the Archaea (Danson *et al.*, 2004; Kerscher *et al.*, 1982; Kerscher and Oesterhelt, 1981; Witzmann and Bisswanger, 1998). OAFORs are simpler and smaller in structure than OADHCs and utilize ferredoxin instead of NAD^+ as physiological electron acceptor (Park *et al.*, 2006; Danson *et al.*, 1984). Based on substrate specificity, OAFORs can be divided into pyruvate ferredoxin oxidoreductase (POR), 2-oxoglutarate ferredoxin oxidoreductase (GOR), and 2-oxoisovalerate ferredoxin oxidoreductase (VOR), which catalyse the same oxidative decarboxylation as the PDHC, OGDHC and BCOADHC, respectively. The importance of OAFORs

has been supported by a recent comparison of the 28 fully sequenced archaeal genomes, which revealed the presence of three to four OAFORs per genome (van Ooyen and Soppa, 2007). Nonetheless, the genes encoding OADHC components were discovered in archaea but the physiological significance of these enzymes has often been questioned (Bryant *et al.*, 2006; McManus, *et al.,* 2006; Kim *et al.,* 2005; Heath *et al.* 2004; Jolley *et al*., 2000). *In vivo* E3 activity has been detected in the halophilic archaea (Danson *et al.*, 1984; Jolley *et al.*, 1996) as well as in *Thermoplasma acidophilum* (Danson *et al.*, 2004 and references therein) but no OADHC activity has so far been reported. The finding of E3 activity in the absence of whole complex activity cannot be construed as clear evidence for the presence of archaeal OADHCs, especially as E3 has other characterized roles that are not associated with the OADHCs (Håkansson and Smith, 2007; Smith *et al.*, 2002 and references therein).

In light of the above, a question emerges as to why *oadhc* genes are found in archaeal genomes. The three *oadhc* operons from *Haloferax volcanii* have sparked interest in the protein evolution of archaeal OADHCs, though their substrate specificity and novel activities still remain unknown. One *oadhc* operon from *H. volcanii* was shown to be transcribed but no *in vivo* OADHC activity could ever be detected (Al Mailem *et al.*, 2008; Jolley *et al.*, 2000). The second *H. volcanii* operon, which only contains genes for the E1α and E1β subunits and an isolated lipoyl domain (Wanner and Soppa, 2002) and the third *H. volcanii oadhc* operon lacking a gene for E3 (van Ooyen and Soppa, 2007), have a role in nitrate-respirative growth, but their exact role and substrate specificity have not been elucidated yet.

More recently, it was shown that the recombinant *Tp. acidophilum* E1, encoded by the $E1_{\alpha}$ and $E1_{\beta}$ genes which are part of a potential *oadhc* operon, had decarboxylase activity with the 2-oxoacids 3-methyl-2-oxopentanoate, 4-methyl-2-oxopentanoate, 3-methyl-2-oxobutanoate and pyruvate (Heath *et al.*, 2004). Analyses of all recombinant proteins encoded in this operon showed that they form an active BCOADHC with an octahedral E2 core (Heath *et al.*, 2007, Caroline Heath, Ph.D. thesis, 2006). From these data a pressing question emerges: If the BCOADHC operon of *Tp. acidophilum* codes for a functional multienzyme

complex, why can no *in vivo* activity be detected? This question becomes even more intriguing in light of a recent proteomics analysis of *Tp. acidophilum* (Sun *et al.*, 2007). This study employing glycerol gradient ultracentrifugation identified a complex of E1 and E2 (> 1 MDa) in *Tp. acidophilum* cells. This complex did not contain E3, although E3 was found in the cells. This is in agreement with reports of E3 activity in *Tp. acidophilum* cell extracts (Smith *et al.*, 1987). Sun *et al.* (2007) do not discuss the absence of the E3 component from the E1-E2 complex. In *B. stearothermophilus* E1 and E3 bind mutually exclusive to E2 and E1 has approximately a two-fold higher affinity for E2 than E3 (Jung *et al.*, 2002 and 2003). The absence of E3 in the purified *Thermoplasma* complex may be explained by the purification process and/or the affinity constants which favour E1 over E3 binding to E2. A biophysical characterization of the *Tp. acidophilum* BCOADHC assembly is addressed in the current thesis (chapter 10).

The presence of *oadhc* genes in archaea could be explained by lateral gene transfer from bacteria (Wanner and Soppa, 2002). Although it is accepted that the *oadhc* genes could be remnants of a gene transfer (reviewed by Brown, 2003), the general notion is that these genes have been retained for a purpose, as it is less likely for organisms with small genomes to accumulate "non-functional" DNA. An alternatively hypothesis is, that the common ancestor of bacteria and archaea possessed *oadhc* genes, and they were only retained in aerobic species (Heath *et al.*, 2004; van Ooyen and Soppa, 2007).

1.4 Lipoylation – a new twist in the story

Recombinant expression of the *Tp. acidophilum* BCOADHC components revealed that yield of active complex was hindered by the inability of the expression host to execute lipoylation of the archaeal E2 under standard growth conditions. Only under specific growth conditions, in which cultures were supplemented with 0.2 mM LA, grown at 30°C for 20 h in darkness and not induced with IPTG, was *E. coli* capable of lipoylating up to 50% of the recombinant E2 (Heath *et al.*, 2007). Hence, attention was drawn to the *Tp. acidophilum* inherent lipoylation system. BLASTP searches using the *E. coli* LipA, LipB and LplA sequences revealed the presence of a putative LplA gene in *Tp. acidophilum*. Since then, two independent groups have been able to crystallize the *Thermo-*

plasma LplA (*Tp.* LplA) (McManus *et al.*, 2006; Kim *et al.*, 2005). Both groups agree that the *Tp.* LplA is a lipoate protein ligase, but whereas McManus *et al.* (2006) could not detect any *in vitro* LplA activity, Kim *et al.* (2005) observed a conversion of LA to lipoyl-AMP, the first catalytic step of LplA, within the crystal structure and concluded from this that the *Thermoplasma* LplA may be catalytically active. Notably, both groups also point out the fact that the *Tp.* LplA is approximately 90 amino acids shorter than most other studied LplAs, possibly missing a C-terminal domain. Further BLASTP searches revealed a gene exactly upstream of the *Thermoplasma* LplA, *ta0513m*, coding for a protein sequence that aligns very well with the C-terminal domain of other known LplAs, but the *ta0513m* gene product has not been studied yet (McManus *et al.*, 2006).

1.5 Aims of the project

It has been known for some time that *Tp. acidophilum* has the *oadhc* genes organized in an operon, which, if recombinantly expressed, yield a functional multienzyme complex. Recently, a proteomics study revealed the presence of the complex components in *Thermoplasma* cell lysate (Sun *et al.*, 2007), but no report of an *in vivo* activity of the complex has yet been shown. With the open questions regarding the activity of the innate *Tp. acidophilum* lipoylation machinery, the aim of this work was to investigate its activity. This involves the cloning, expression, purification and characterisation of the lipoate protein ligase as well as a second protein encoded by *ta0513m,* which has been indicated to have a possible role in lipoylation. Moreover, the assembly of the *Thermoplasma* BCOADHC will be studied to confirm if its complex architecture follows a certain stoichiometry like eukaryotic and prokaryotic OADHCs.

CHAPTER 2
MATERIALS

2.1 Chemicals

Chemical	Chemical Supplier
Acetonitrile (HPLC grade)	Fisher Scientific Ltd., Loughborough, UK
Agarose	Sigma-Aldrich Company Ltd., Gillingham, UK
Ammonium sulphate	Sigma-Aldrich Company Ltd., Gillingham, UK
Ampicillin	Melford Laboratories Ltd., Suffolk, UK
Bovine serum albumin (BSA)	New England Biolabs Ltd., Hitchin, UK
Bromophenol blue	Sigma-Aldrich Company Ltd., Gillingham, UK
Carbenicillin	Sigma-Aldrich Company Ltd., Gillingham, UK
CHAPS	Melford Laboratories Ltd., Suffolk, UK
Chloramphenicol	Sigma-Aldrich Company Ltd., Gillingham, UK
Dithiothreitol (DTT)	Melford Laboratories Ltd., Suffolk, UK
dNTPs	Bioline Ltd. (London, UK)
EDTA	Sigma-Aldrich Company Ltd., Gillingham, UK
Ethanol (AR grade)	Fisher Scientific Ltd., Loughborough, UK
Ethidium bromide	Sigma-Aldrich Company Ltd., Gillingham, UK
Glacial acetic acid	Fisher Scientific Ltd., Loughborough, UK
Glycerol	Fisher Scientific Ltd., Loughborough, UK
Isopropyl-1-thio-ß-galactopyranoside (IPTG)	Melford Laboratories Ltd., Suffolk, UK
Kanamycin sulphate	Melford Laboratories Ltd, Suffolk, UK
D,L Lipoic acid	Sigma-Aldrich Company Ltd., Gillingham, UK
Methanol	Fisher Scientific Ltd., Loughborough, UK
2-(N-Morpholino)ethanesulfonic acid (MES)	Melford Laboratories Ltd., Suffolk, UK
3-(N-Morpholino)propanesulfonic acid (MOPS)	Melford Laboratories Ltd., Suffolk, UK
Sodium chloride	Sigma-Aldrich Company Ltd., Gillingham, UK
Sodium dodecylsulphate (SDS)	Sigma-Aldrich Company Ltd., Gillingham, UK
Trifluoroacetic acid (TFA) (Certified)	Fisher Scientific Ltd., Loughborough, UK
Tris base	Melford Laboratories Ltd., Suffolk, UK
Trypton	Melford Laboratories Ltd., Suffolk, UK
Yeast Extract	Melford Laboratories Ltd., Suffolk, UK

Triton-X100	Fisher Scientific Ltd., Loughborough, UK
Tween-20	Sigma-Aldrich Company Ltd., Gillingham, UK
Helmanex II solution	Hellma GmbH & Co, Müllheim, Germany

2.2 Molecular biology reagents

Reagent	**Reagent Supplier**
Phusion polymerase	New England Biolabs Ltd., Hitchin, UK
E. coli BL21(DE3)	Novagen-Merck, Nottingham, UK
E. coli Rosetta(DE3)	Novagen-Merck, Nottingham, UK
E. coli ArcticExpress™(DE3)	Stratagene Europe, Netherlands
E. coli ArcticExpress™(DE3)RIL	Stratagene Europe, Netherlands
Restriction digest enzymes	New England Biolabs Ltd., Hitchin, UK Promega, Southampton, UK
Blue/orange ready-to-use 6x loading dye	Promega, Southampton, UK
E. coli JM109 cells	Promega, Southampton, UK
pGEM-T Easy vector	Promega, Southampton, UK
Primers	MWG Biotech, Ebersberg, Germany
T4 DNA ligase	Promega, Southampton, UK
Taq polymerase	Genesys, Camberley, UK
QIAEX II Gel Extraction Kit	Qiagen, Crawley, UK
QIAprep® Spin Miniprep Kit	Qiagen, Crawley, UK
PCR purification kit	Qiagen, Crawley, UK
pET expression vectors	Novagen-Merck, Nottingham, UK
Hyperladder I and IV	Bioline Ltd., London, UK
PageRuler™ Unstained Protein Ladder	Fermentas, York, UK

2.3 Protein purification reagents

Reagent	**Reagent Supplier**
Bis-Tris and Tris-Glycine precast gels	Invitrogen Ltd., Paisley, UK
HBS-EP buffer	GE Healthcare UK Ltd., Bucks., UK
Heparing Sepharose Fast Flow, 5 ml	GE Healthcare UK Ltd., Bucks., UK
HiLoad 16/10 Q Sepharose High Performance 20 ml column	Amersham Pharmacia

HisTrap FF, HisTrap HP chromatography columns	GE Healthcare UK Ltd., Bucks., UK
NSB reducer	GE Healthcare UK Ltd., Bucks., UK
NuPAGE electrophoresis system	Invitrogen Ltd., Paisley, UK
Protein standards	Bio-Rad (Hemel Hempstead, UK) / Invitrogen / GE Healthcare
HiTrap Q FF, ionexchange chromatography columns	GE Healthcare UK Ltd., Bucks., UK
Resource Q column	GE Healthcare UK Ltd., Bucks., UK
Sensor CM5 chips	GE Healthcare UK Ltd., Bucks., UK
PD10 columns	GE Healthcare UK Ltd., Bucks., UK
Silver stain kit	Fermentas, York, UK
Superdex 200 10/300 column	Amersham Pharmacia
Bis-Tris and Tris-Glycine precast gels	Invitrogen Ltd., Paisley, UK
Bradford assay reagent	Bio-Rad, Hemel Hempstead, UK
Heparin Sepharose FF	Amersham Pharmacia
HiLoad 16/10 Q Sepharose High Performance	Amersham Pharmacia
HiLoad 16:60 Superdex 75 prep grade	Amersham Pharmacia
NuPAGE electrophoresis system	Invitrogen Ltd., Paisley, UK
Pre-stained Protein standards	Invitrogen Ltd., Paisley, UK
SnakeSkin™ dialysis tubing	Pierce, Thermofisher Scientific Inc., UK
Whatman® Anotop 10, 0.02 µm filter	Fisher Scientific Ltd., Loughborough, UK
Anti-lipoic acid rabbit antibody	Calbiochemm, Nottingham, UK
Anti-rabbit goat IgG antibody	Sigma-Aldrich Company Ltd., Gillingham, UK

2.4 Protein crystallography reagents

Reagent	**Supplier**
High vacuum grease	Dow Corning Corp. U.S.A
Siliconized circle cover slides 22 mm	Hampton Research, U.S.A
Structure Screen 1	Molecular dimensions Ltd., Suffolk, UK
Structure Screen 2	Molecular dimensions Ltd., Suffolk, UK
Tacsimate	Hampton Research, U.S.A

CHAPTER 3
General methods and protein purification

3.1 DNA methods

3.1.1 Polymerase chain reaction (PCR)

Thermoplasma acidophilum genes were PCR amplified from *Tp. acidophilum* genomic DNA (100 µg/µl), kindly provided by Dr Alex Jefferies, University of Bath. PCR reactions with a final volume of 50 µl contained template DNA (100 ng), forward and reverse primers (0.5 µmol each), Phusion DNA polymerase (1U) and 200 µM of each deoxynucleoside triphosphate (dATP, dCTP, dGTP, dTTP) in Phusion HF buffer supplied by the manufacturer (New England Biolabs). Controls involved reactions that contained only a single primer or no template DNA.

Reactions were carried out in an Eppendorf Mastercycler (Eppendorf, Germany). A standard cycle comprised:

	Hot start denaturation	(98°C)	30 s
30 cycles	Denaturation	(98°C)	10 s
	Annealing	(X°C)	30 s
	Extension	(72°C)	30 s
	Final extension	(72°C)	10 min

X is the temperature ~ 5°C below the melting temperature (T_m) of the primers.

3.1.2 Agarose gel electrophoresis

DNA was visualised by gel electrophoresis in 0.8 – 1% (w/v) agarose supplemented with 0.5 µg/ml ethidium bromide, made by heat-dissolving in TAE buffer [40 mM Tris acetate, pH 8.0, 1 mM EDTA]. DNA samples were prepared for electrophoresis by the addition of an appropriate amount of ready-to-load 6x loading buffer [0.4% orange G, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 15% Ficoll® 400, 10 mM Tris-HCl (pH 7.5) and 50 mM EDTA (pH 8.0)]. DNA was electrophoresed in TAE buffer at a constant 60-80 V. Gels were run for an amount of time that facilitated adequate band separation; progress of DNA through the gel was tracked by following the xylene cyanol dye front.

Ethidium bromide-bound DNA bands were visualised under a UV transilluminator. The size of fragments was estimated by calibration against DNA molecular weight marker run simultaneously on the gel.

3.1.3 DNA gel purification

DNA bands were excised from agarose gels and then purified using the QIAEX II Gel Extraction Kit, following the manufacturer's protocol (Qiagen, 1999). DNA was eluted in 20 µl of 10 mM Tris-HCl (pH 8.5) and stored at -20°C. Concentration of eluted DNA was estimated by electrophoresis of a defined amount of sample in an agarose gel (as in section 3.1.2) and comparison with a DNA mass ladder (Bioline), run in the gel simultaneously

3.1.4 Growth of *Escherichia coli*

Escherichia coli strains were cultured in liquid Luria Bertani (LB) medium adjusted to pH 7.0 with NaOH, containing sodium chloride (0.5% w/v), tryptone (1.0% w/v) and yeast extract (0.5% w/v); for LB plates, agar (1.5% w/v) was added to the above medium. The medium was supplemented with antibiotic(s) when required (ampicillin, 100µg/ml; carbenicillin, 50 µg/ml; kanamycin, 30 µg/ml; chloramphenicol, 34 µg/ml) as recommended in the Novagen pET System manual (11^{th} edition) for pET vectors, or the Promega manual (1999) for pGEM-T vectors.

3.1.5 Growth of *Tp. acidophilum* cell cultures

Tp. acidophilum (strain 1728) was purchased from DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Braunschweig, Germany). Cell cultures were started from the vacuum-dried cell pellets following the supplier's instructions. The growth medium composition is given in Table 1. To examine the effect of lipoic acid (LA), cell cultures were supplemented with 0.2 mM LA. The medium was adjusted to pH 1.0 – 2.0 with 10 M H_2SO_4. The starter culture was subcultured five times with and without LA supplement.

Table 1 *Tp. acidophilum growth* medium composition

1L DSM 158 medium contains		**1L trace element solution in distilled H_2O contains**	
$(NH_4)_2SO_4$	1.32 g	$FeCl_3.6\ H_2O$	1.93 g
KH_2PO_4	0.372 g	$MnCl_2.4\ H_2O$	0.18 g
$MgSO_4.7\ H_2O$	0.247 g	$Na_2B_4O_7.10\ H_2O$	0.45 g
$CaCl_2.2\ H_2O$	0.074 g	$ZnSO_4.7\ H_2O$	22.0 mg
Yeast extract (Difco, stock solution 10% w/v)	1.0 g	$CuCl_2.2\ H_2O$	5.0 mg
D-glucose (stock solution 50% w/v)	10.0 g	$Na_2MoO_4.2\ H_2O$	3.0 mg
Trace element solution	10 ml	$VOSO_4.5\ H_2O$	3.8 mg
		$CoSO_4.7\ H_2O$	2.0 mg

3.1.6 Plasmid DNA purification

Cells were grown overnight at 37°C in 5 – 10 ml LB medium (section 3.1.4) with shaking. Cells were harvested by centrifugation at 6000 × g for 6 min. Plasmid DNA was extracted using the QIAprep® Spin Miniprep Kit following the manufacturer's protocol (Qiagen, 2005). Plasmids could then be visualised by electrophoresis in a 1.0% agarose gel (section 3.1.2).

3.1.7 A-tailing of blunt-ended PCR products

Gel-purified PCR fragments generated by Phusion DNA polymerase were A-tailed using *Taq* DNA polymerase, following the Promega A-tailing procedure protocol (pGEM-T Vector Systems Technical Manual, 1999). The reaction was carried out in 0.5 ml PCR tubes (with low DNA binding capacity) at 70°C for 30 min.

3.1.8 Sequencing of PCR products

PCR products were A-tailed (section 3.1.5), ligated into pGEM-T Easy vector (section 3.1.9) and plated onto agar plates as per the manufacturer's instructions (Promega). The multiple cloning region of the pGEM-T Easy vector falls

within the α-peptide coding region for β-galactosidase. If grown on agar supplemented with 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (X-Gal), the β-galactosidase activity is detected by the blue colour of colonies. Successful insertion of PCR products into the pGEM-T Easy multiple cloning site abolishes β-galactosidase activity and hence recombinant clones will form white colonies. From each ligation reaction, five white colonies with the potentially correct DNA insert and one blue colony were selected, inoculated in 10 ml LB medium and incubated overnight at 37°C. From the liquid overnight cultures pGEM-T Easy plasmid DNA was purified as described in section 3.1.6. Plasmid DNA (300 – 500 ng) was supplied with Sp6 and/or T7 primer and sent for sequencing at the University of Bath or to Geneservice Ltd. (Cambridge, U.K).

3.1.9 Cloning: restriction enzyme digests and DNA ligations

A typical restriction digest contained ≥ 200 ng of plasmid DNA and was performed in 1.5 ml Eppendorf tubes in total volumes that ranged from 10 – 100 µl. The number of units of restriction enzyme(s) used depended upon the amount of DNA being cut, and was usually 10 – 20 U. For double digests, a buffer that was compatible for both restriction enzymes was used (details of compatibility were obtained from the manufacturer's catalogue). If enzymes were not compatible, a sequential digest was performed. BSA (100 µg/ml) was added only when recommended for an enzyme. Glycerol concentrations were kept below 10% (v/v) to avoid star activity. Reactions were usually incubated for 1 to 3 h at 37°C. After incubation, samples were either stored at -20°C, or run on an agarose gel. Target DNA fragments were excised from the gel and purified using the QIAEX II gel extraction kit (section 3.1.3). 'Single restriction digestion' controls were reactions containing only a single restriction enzyme.

Target genes were ligated into cut plasmid vectors using the Promega Rapid Ligation Kit following the accompanying protocol. Reactions (10 µl total volume) were carried out in 0.5 ml PCR tubes (with low DNA binding capacity). Insert:vector molar ratios were always calculated according to the Promega Rapid Ligation protocol. In the negative ligation control, no insert DNA was supplied in the mixture, whereas the positive ligation control included control insert DNA provided by the supplier. Reaction mixtures were incubated at 4°C overnight.

3.1.10 Cloning of the *Thermoplasma acidophilum* lipoate protein ligase (LplA)

The *Tp. acidophilum* lipoate protein ligase (*Tp.* LplA) is encoded by the gene *ta0514* (see chapter 4). The forward and reverse primers for the *ta0514* gene were

5'-CAT**ATG**GAAGGCAGGCTTCTTTTACTAGAAACA-3' (start codon in bold and NdeI restriction site underlined) and

5'-CGGGATCC**CTA**TACGACCTCTTTCCTCAG-3' (stop codon in bold and BamHI restriction site underlined) respectively. PCR amplification of *Tp.* LplA from genomic DNA gave one PCR product of the expected size (Fig. 3.1).

The PCR product was ligated (section 3.1.9) into pGEM-T Easy vector and sent for sequencing. Correct clones were ligated into pET19b via the NdeI and BamHI restriction sites to allow the expression of recombinant N-terminal His-tagged LplA.

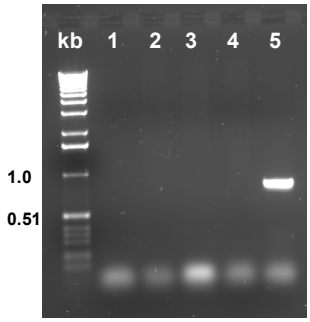


Fig. 3.1 PCR amplification of the *Tp. acidophilum* LplA DNA sequence (1% agarose gel). Marker sizes are indicated in kilo base pairs. Lane 1, no DNA; lane 2, no reverse primer; lane 3, no forward primer; lane 4, PCR amplification reaction without DMSO; lane 5, PCR amplification reaction with DMSO.

3.1.11 Cloning of the putative *Thermoplasma acidophilum* C-terminal domain (CTD) of LplA

The *Thermoplasma* CTD is encoded by the gene *ta0513m* (see chapter 4). The forward and reverse primer were

5'-GCTAGC**ATG**CATATGATGTACAGCAAGAACTGGAAG-3' (start codon bold and NheI restriction site underlined) and

5'-GGATCC**TCA**GATCACCCTCAAAGCCTGAATAAAGTC-3' (stop codon in bold and BamHI site underlined) respectively. The PCR reaction gave rise to

one product (Fig. 3.2). The PCR product was ligated (section 3.1.9) into pGEM-T Easy vector and sent for sequencing. The correct clone was ligated into pET28a for expression of recombinant CTD with an N-terminal His-tag.

For expression of a nontagged version of the CTD, the pET28a plasmid DNA containing the putative *ctd* gene was sequentially digested with NheI and BamHI. The CTD gene insert was gel purified and ligated into linearized pET24 cut with the same restriction enzymes. The pET24-CTD plasmid was transformed into BL21(DE3) *E. coli* cells and selected for the appropriate antibiotic resistance. Transformants carrying the correct insert were confirmed by sequencing.

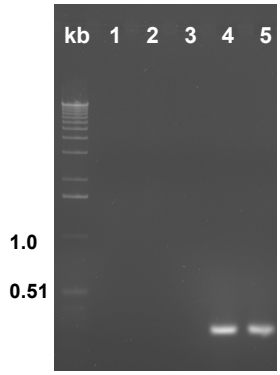


Fig. 3.2 PCR amplification of *Tp. acidophilum* CTD DNA sequence (1% agarose gel). Marker sizes are indicated in kilo base pairs. Lane 1, no DNA; lane 2, no reverse primer; lane 3, no forward primer; lane 4, PCR amplification reaction without DMSO; lane 5, PCR amplification reaction with DMSO.

3.1.12 Cloning of the *Thermoplasma acidophilum* E2 lipoyl domain (E2lipD)

The dihydrolipoyl acyl-transferase (E2) is encoded by the *Tp. acidophilum* gene *ta1436*. The E2 component is composed of three domains: the E2 lipoyl domain (E2lipD), the peripheral subunit binding domain (PSBD), with which either E1 or E3 associates, and the catalytic domain. To clone only the *Thermoplasma* E2lipD, the forward primer was designed to match the *ta1436* gene and the reverse primer was designed to match the gene sequence coding for the C-terminus of the E2lipD and part of the linker region between the lipoyl and peripheral subunit binding domains (Fig. 3.3).

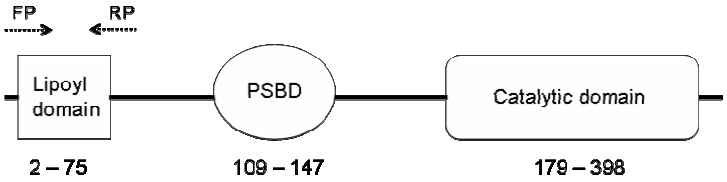


Fig. 3.3 The domain structure of the *Tp. acidophilum* E2 as predicted from a Pfam A search. The Pfam A search revealed three domains. The start and end of the domain is indicated by the amino acid number below. The forward primer (FP) and reverse primer (RP) are illustrated.

The reverse primer was designed to introduce a stop codon at the C-terminal end of the E2lipD. The sequences for the forward and reverse primer were 5'-CGCCAT**ATG**TACGAATTCAAACTGCCAGACATAGG-3' (start codon in bold and NdeI restriction site underlined) and 5'-CGGGATCC**TCA**GGTATCGATCTGCAGGAG-3' (stop codon in bold and BamHI restriction site underlined) respectively. The PCR reaction gave one single product of the expected size (Fig. 3.4) that was ligated (section 3.1.9) into pGEM-T Easy vector and sent for sequencing. A correct clone was ligated into pET19b vector cut with NdeI and BamHI for expression of recombinant E2lipD carrying an N-terminal His-tag.

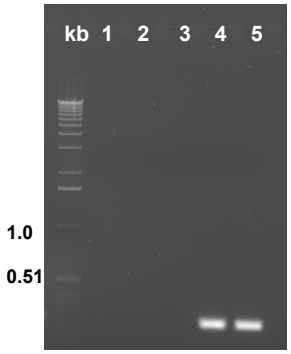


Fig. 3.4 PCR amplification of *Tp. acidophilum* E2 lipoyl domain DNA sequence (1% agarose gel). Marker sizes are indicated in kilo base pairs. Lane 1, no DNA; lane 2, no reverse primer; lane 3, no forward primer; lane 4, PCR amplification reaction without DMSO; lane 5, PCR amplification reaction with DMSO.

3.1.13 Transformation by heat-shock

All transformations were done using commercially available competent cells of various *E. coli* strains. Each transformation reaction contained approximately 1

ng plasmid DNA to which 10 µl of competent cells were added. The incubation time of competent cells with the DNA on ice, as well as the duration of heat shock in a water bath at 42°C, varied between 5 to 20 min, and 30 to 50 s, respectively based on the cell line and the manufacturer's recommendations. After the heat shock, 500 µl room-temperature LB medium was added, and cells were incubated at 37°C for 1.5 – 2 h with shaking. Cells were spread onto LB agar plates containing the appropriate antibiotic. For blue-white selection of transformants containing ligated pGEM-T clones, cells were spread onto LB agar plates as per the manufacturer's instructions.

3.2 Protein methods

3.2.1 Non-denaturing and SDS-PAGE

Non-denaturing PAGE was prepared and carried out as described below, omitting SDS and β-mercaptoethanol in any preparations; also, samples were not boiled. Pre-cast NuPAGE® gels (Invitrogen) were run according to the manufacturer's instructions.

Gel preparation: A resolving gel (15%) was made by mixing 4 ml of 30% (v/v) acrylamide, 2.6 ml buffer [1.0 M Tris-HCl, pH 8.8], 80 µl of 10% (w/v) SDS, and distilled water to a final volume of 7.98 ml, and was polymerised by the addition of 12 µl of 20% (w/v) ammonium persulphate and 10 µl tetramethylethylenediamine (TEMED). The stacking gel (4% acrylamide) was made by mixing 0.625 ml of 30% (v/v) acrylamide, 0.625 ml buffer [1.0 M Tris-HCl, pH 6.8], 50 µl of 10% (w/v) SDS, and 3.6 ml distilled water, and was polymerised with the addition of 15 µl of 20% (w/v) ammonium persulphate and 5 µl TEMED.

Preparation of protein samples: Samples were prepared for loading by the addition of an equal volume of 2 x loading buffer [0.05 M Tris-HCl, pH 6.8, 4% (w/v) SDS, 20% (w/v) sucrose, 10% (v/v) β-mercaptoethanol, 0.2% (w/v) bromophenol blue]. Samples were heated at 100°C for ~ 3 min and then vortexed vigorously. Samples were either loaded immediately onto gels, or stored at -20°C until use.

Electrophoresis: Electrophoresis buffer was composed of 0.12 M Tris base, 9.4% (w/v) glycine, and 0.5% (w/v) SDS. The gel was stained in Coomassie

staining solution [0.25% (w/v) Coomassie Brilliant Blue R250, 45% (v/v) methanol, 45% (v/v) deionised water, 10% (v/v) acetic acid]. Destain solution was 20% (v/v) methanol, 10% (v/v) acetic acid and 70% (v/v) distilled water.

3.2.2 Optimization of chromatography conditions for protein purification

The chromatography methods used for protein purification include affinity purification, ion exchange, and gel permeation chromatography. Optimum pH and purification resins were tested using the Äkta FPLC purifier scouting settings (Amersham Bioscience). Chromatography columns used include Resource Q, HiTrap Q FF, Q Sepharose, Resource S and Heparin Sepharose. The pH range tested was pH 6.0 – 9.0 as well as pH 4.5 – 6.0. The purification conditions for each protein given in this thesis are based on optimization trials.

3.2.3 Buffer exchange of protein samples

Proteins were buffer exchanged by either dialysis or by using PD10 desalting columns. SnakeSkin^{TM} dialysis tubing (Pierce) was prepared according to the manufacturer's instructions. Samples were sealed in the dialysis tubing, and the tubing was submerged in the desired buffer with a sample to buffer volume ratio of 1:25, with at least three changes, at 4°C with gentle stirring

PD10 desalting columns (GE Healthcare) were used to exchange the His-purification elution buffer for the final storage buffer following the manufacturer's instructions.

3.2.4 Concentrating purified protein samples

Stirred-tank concentrators (Amicon) were used to concentrate large volumes of purified proteins. Smaller volumes of dilute protein samples were concentrated at 4°C using Amicon Ultra-15 or Amicon Ultra-4 centrifugal filter devices (Millipore) according to the manufacturer's instructions.

3.2.5 Expression and purification of *Tp. acidophilum* 2-oxoacid decarboxylase (E1)

The E1α and E1β genes were available pre-cloned in pET19b and pET28a, respectively, by Dr. Caroline Heath, University of Bath (Heath *et al.*, 2004). The

recombinant proteins carry N-terminal His-tags. For expression, bacterial cell cultures were grown at 37°C in LB medium supplemented with the appropriate antibiotics to an OD_{600nm} of 0.6 and induced with 1 mM IPTG for 4 h with shaking. Cells were harvested by centrifuging at 6000 x g for 10 min at 4°C. The harvested cells were sonicated in the running buffer (20 mM sodium phosphate, pH 7.4, 0.5 M NaCl) and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography using a 1 ml HisTrap HP column applying a linear gradient of imidazole (0 – 0.5 M) in the running buffer. Fractions containing the recombinant E1αβ were pooled and concentrated in Amicon ultra (10 MWC) concentrator devices (Millipore) as per the manufacturer's instructions. The concentrated samples were desalted using a PD10 column (GE Healthcare) (section 3.2.3) by eluting the protein with 20 mM sodium phosphate, pH 7.4.

3.2.6 Expression and purification of *Tp. acidophilum* dihydrolipoyl acyltransferase (E2)

The E2 gene was available pre-cloned in pET28a by Dr. Caroline Heath, University of Bath (Heath *et al.*, 2007). The recombinant protein carries an N-terminal His-tag. For expression of unlipoylated E2, cells were induced with 1 mM IPTG at 37°C and grown for 4 h. Soluble lipoylated E2 protein (50% lipoylation) was obtained by expressing the E2 gene in uninduced BL21(DE3) cells grown in LB medium supplemented with lipoic acid (0.2 mM), at 30°C for 20 h. Harvested cells were resuspended in the running buffer (20 mM sodium phosphate, pH 7.4, 0.5 M NaCl), sonicated and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on a 1 ml HisTrap HP column applying a linear gradient of imidazole (0 – 0.5 M) in the running buffer. Fractions containing E2 were concentrated using Amicon ultra (10 MWC) concentrator devices (Millipore). Concentrated E2 was buffer exchanged (section 3.2.3) into 20 mM sodium phosphate, pH 7.4.

3.2.7 Expression and purification of *Tp. acidophilum* dihydrolipoamide dehydrogenase (E3)

The E3 gene had been pre-cloned into pET28a and expressed in *E. coli* BL21(DE3) cells (Heath *et al.*, 2007). The recombinant protein carries an N-

terminal His-tag. Cultures were inoculated from glycerol stocks. Cells were not induced with IPTG. After an OD_{600nm} of ~ 0.6 was reached, expression was allowed for 5 h at 37°C with shaking. Harvested cells were resuspended and sonicated in the running buffer (20 mM sodium phosphate buffer, pH 7.4, 0.5 M NaCl, 20 mM imidazole). The cell lysate was centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on a HisTrap HP column applying a linear gradient of imidazole (0 – 2 M) in 20 the running buffer. The protein was buffer exchanged (section 3.2.3) into 20 mM sodium phosphate, pH 7.4.

3.2.8 Expression and purification of His-tagged *Tp. acidophilum* LplA (His-LplA)

The pET19b carrying the *lpla* gene (section 3.1.10) was transformed into ArcticExpress(DE3) *E. coli* cells (Stratagene). Cell cultures were grown in LB medium at 37°C using shake flasks until an OD_{600nm} of 0.7 – 0.9 was reached. Recombinant protein was induced with a final concentration of 0.25 mM IPTG and incubated for 3 h at 12°C with shaking. The cells were lysed by sonication in running buffer (20 mM Tris-HCl, pH 7.5, 0.3 M NaCl, 20 mM imidazole) in the presence of 2% (v/v) Triton-X 100. The lysate was centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on a HisTrap FF column (GE Healthcare) applying a linear gradient of imidazole (0.02M – 0.5 M) in the running buffer. Fractions containing His-LplA from the first Ni^{2+} affinity chromatography purification step were pooled, diluted 5-fold in running buffer, and re-applied to a 1 ml HisTrap HP column (GE Healthcare) for a second cycle of Ni^{2+} affinity chromatography. Fractions from the second Ni^{2+} affinity chromatography were again pooled, diluted 10-fold in 50 mM Tris-HCl, pH 8.0, and subjected to ion-exchange chromatography using a 1 ml HiTrap Q FF column and eluted by applying a linear gradient of NaCl (0 – 0.5 M) in 50 mM Tris-HCl, pH 8.0. The protein purification steps are illustrated in Fig. 3.2. Protein fractions, > 90% pure (by analysis on gel), were collected and stored in elution buffer at -20°C until further use.

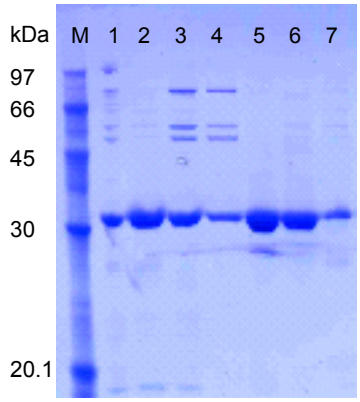


Fig. 3.2 SDS-PAGE gel showing the different stages of purification of *Tp. acidophilum* His-tagged LplA. His-LplA was purified from a 0.5 L culture grown as described above. M, protein marker (kDa); lane 1, LplA from the initial Ni^{2+} affinity purification; lanes 2 – 4, fractions of purified LplA from a second Ni^{2+} affinity purification; lanes 5 – 7, fractions of purified LplA after anion exchange on Q FF.

3.2.9 Expression and purification of *Tp. acidophilum* nontagged LplA (ntLplA)

The pET3a plasmid carrying the *Thermoplasma lpla* gene was a kind gift from Edward McManus of Prof. Richard Perham's group, University of Cambridge. Details of the construct are described in McManus *et al.* (2007).

The pET3a-LplA plasmid was transformed into BL21(DE3) *E. coli* cells following the supplier's instructions. Cultures were grown to an OD_{600nm} 0.7 – 0.9 and induced with 0.5 mM IPTG at 37°C for 3 – 4 h. Cells were harvested, sonicated in 20 mM Tris-HCl, pH 8.5, 0.5% (v/v) Triton X-100, and centrifuged at 75,000 x g at 4°C for 20 min. The supernatant was subjected to anion exchange on Hi-Trap Q FF applying a linear gradient of NaCl (0 – 0.5 M) in 20 mM Tris-HCl, pH 8.5, which yielded ~ 60 % pure ntLplA (by analysis on gel). Fractions containing the ntLplA protein were diluted 10 fold in 20 mM Tris-HCl, pH 8.5, and the anion exchange purification step was repeated, yielding ≥ 80 % pure ntLplA (by analysis on gel). An overview of the protein purification is shown in Fig. 3.3. The purified protein was stored in the elution buffer at -20°C until further use.

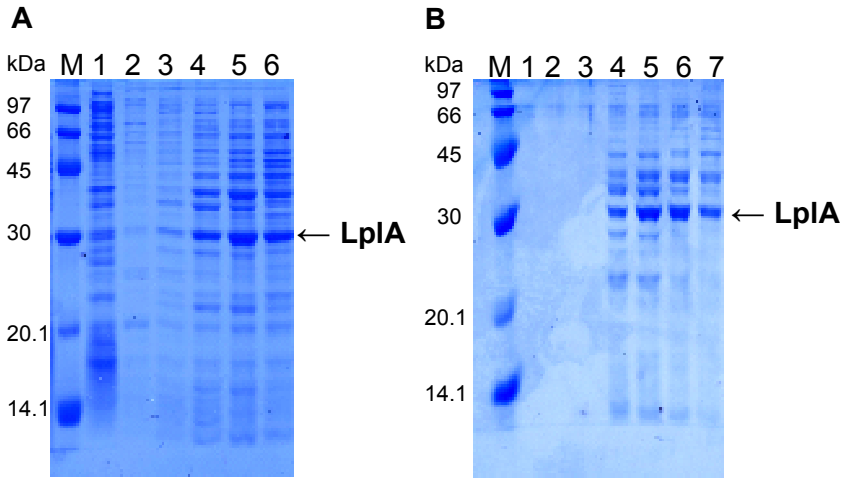


Fig. 3.3 SDS-PAGE gels showing the different stages of purification of non-tagged LplA. M, protein marker (kDa). **A**: Initial anion exchange chromatography on HiTrap Q FF. Lane 1, flow-through from Q FF column; lanes 2 – 6, fractions eluted with increasing NaCl gradient; lanes 4 – 6, fractions of purified ntLplA. **B**: Second anion exchange chromatography of ntLplA. Lane 1, flow through from Q FF column; lanes 2 – 6, fractions eluted with increasing NaCl gradient; lane 7, ntLplA, concentrated using Amicon ultra spin filter devices.

3.2.10 Expression and purfication of *Tp. acidophilum* His-tagged CTD (His-CTD)

The pET28a plasmid carrying the *ctd* gene (section 3.1.11) was transformed into ArcticExpress(DE3) *E. coli* cells (Stratagene) following the manufacturer's instructions. Cell cultures were grown at 37°C in LB medium using shake flasks. Recombinant protein was induced at OD_{600nm} of 0.7 – 0.9 with 0.25 mM IPTG for 4 h at 12°C with shaking. Cells were sonicated in running buffer (20 mM Tris-HCl, pH 8.5, 0.3 M NaCl, 20 mM imidazole) with 2% (v/v) Triton X-100 and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on a HisTrap FF column, applying a linear imidazole gradient (0 – 0.5 M) in the running buffer. Initial Ni^{2+} affinity chromatography yielded ≥ 85% pure His-CTD (by analysis on gel). Fractions containing His-CTD from the first Ni^{2+} affinity chromatography purification step were pooled, diluted 5-fold in running buffer, and re-applied to a 1 ml HisTrap HP column (GE Healthcare) for a second cycle of Ni^{2+} affinity chromatography, yielding ≥ 95% pure protein (by analysis on gel). An overview of the protein purification is shown in Fig. 3.4. Protein fractions were pooled and stored in elution buffer with 5% (v/v) glycerol at -20°C until further use.

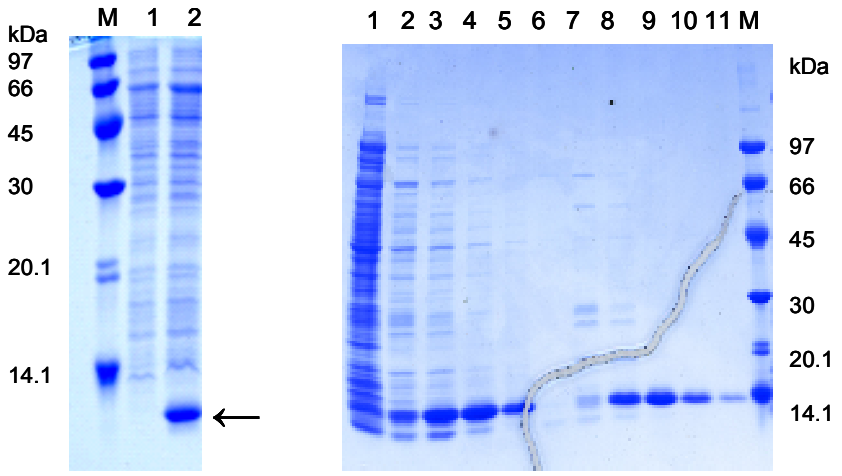


Fig. 3.4 SDS-PAGE gels showing the purification steps of His-tagged CTD. M, protein marker (kDa). **A**: Induction of recombinant His-CTD in ArcticExpress *E. coli* cells. Lane 1, uninduced cells; lane 2, after 4 h induction. His-CTD indicated by arrow. **B**: Consecutive Ni^{2+} affinity purification using a HisTrap FF column. Lanes 1 – 5, initial Ni^{2+} affinity purification, lanes 6 – 11 second Ni^{2+} affinity purification. Lanes 1 and 6, flow through from the Ni^{2+} column.

3.2.11 Expression and purification of *Thermoplasma* nontagged CTD (ntCTD)

The pET24a plasmid carrying the *ctd* gene (section 3.1.11) was transformed into ArcticExpress(RIL) *E. coli* cells (Stratagene) following the manufacturer's instructions. Cell cultures were grown at 37°C in LB medium until OD_{600nm} of 0.7 – 0.9 and induced with 0.25 mM IPTG for 3 – 4 h at 12°C with shaking. Cells were harvested, sonicated in 20 mM Tris-HCl, pH 8.5, and centrifuged at 75,000 x g at 4°C for 20 min. To enrich ntCTD before ion exchange chromatography, the cell lysate was subjected to ammonium sulphate precipitation using 1.5 M and 2.5 M $(NH_4)_2SO_4$ as per protocol ([Online] available at http://proteincrystallography.org/protein-purification/analytical-ammonium-sulphate-cut.php, accessed December 16th, 2008). The majority of the recombinant ntCTD was in the 2.5 M $(NH_4)_2SO_4$ pellet and was buffer exchanged (section 3.2.3) into 50 mM Tris-HCl, pH 7.5. The resulting supernatant was subjected to anion exchange chromatography on HiTrap Q FF and the protein was eluted with a linear gradient of NaCl (0 – 0.5 M) in 20 mM Tris-HCl, pH 7.5 (Fig. 3.5).

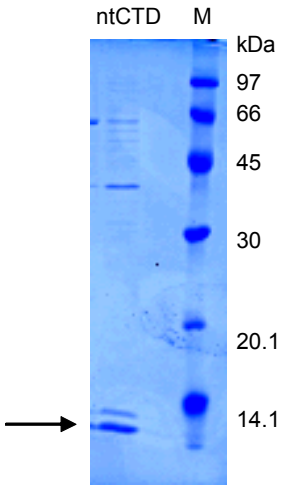


Fig. 3.5 SDS-PAGE gel of purified ntCTD after ammonium sulphate precipitation and ion exchange chromatography. M, protein marker (kDa). Purification of ntCTD was carried out as described (section 3.2.11). ntCTD indicated by arrow.

3.2.12 Expression and purification of *Thermoplasma* E2 lipoyl domain

The pET19b plasmid carrying the *e2lipd* gene (section 3.1.12) was transformed into Rosetta(DE3) *E. coli* cells following the supplier's protocol and cell cultures were grown in LB medium at 37°C. At OD_{600nm} of 0.6 – 0.9 cells were induced with 0.5 mM IPTG and grown for ~ 3 h at 37°C with shaking. Cells were sonicated in running buffer (20 mM Tris-HCl, pH 8.0, 0.3 M NaCl, 20 mM imidazole) and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on HisTrap FF applying a linear gradient of imidazole (0 – 0.5 M) in 20 mM Tris-HCl, pH 8.0, 0.3 M NaCl (Fig. 3.6). Fractions containing the His-tagged E2lipD were pooled, diluted 10-fold in 50 mM Tris-HCl, pH 8.0, and reapplied to a HiTrap Q FF column for ion exchange chromatography applying a linear gradient of NaCl (0 – 0.5 M) in 50 mM Tris-HCl, pH 8.0. Purified E2lipD (≥ 90% pure by analysis on gel) was stored in elution buffer with 10% (v/v) glycerol at -20°C until further use.

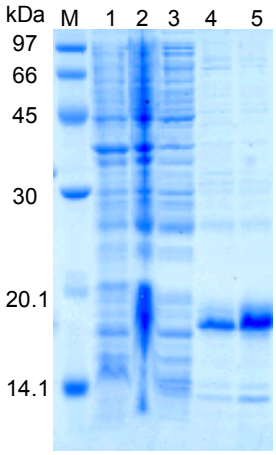


Fig. 3.6 SDS-PAGE gel of initial Ni^{2+} affinity purification of E2lipD. M, protein marker (kDa). Lane 1, uninduced E2lipD; lane 2, induced E2lipD; lane 3, flow through from the Ni^{2+} column; lanes 4 and 5, fractions containing E2lipD, eluted from Ni^{2+} column.

3.2.13 Expression and purification of the *Escherichia coli* LplA

The TM202 plasmid carrying the *E. coli* LplA gene was a kind gift from Prof. R. Perham (University of Cambridge), and was transformed into BL21(DE3) cells as per the instructions of the supplier. Cultures were grown in LB and recombinant protein was induced with 0.5 mM IPTG at OD_{600nm} of 0.6 – 08 for ~ 3 h at 37°C. Cells were sonicated in 50 mM Tris-HCl, pH 8.5, and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to anion exchange chromatography on HiTrap Q FF and pure protein was eluted by applying a linear NaCl gradient (0 – 0.5 M) in 50 mM Tris-HCl, pH 8.5. Fractions containing purified protein were concentrated, loaded onto a SuperdexTM gel filtration column (10/300GL) and eluted using isocratic conditions of 20 mM Tris-HCl, pH 7.5, 300 mM NaCl and stored in elution buffer with 5% (v/v) glycerol at -20°C until further use.

3.2.14 Expression and purification of the *Escherichia coli* E2 lipoyl domain

The pET11c over-expression plasmid carrying the *E. coli* lipoyl domain was a gift from Prof. R. Perham (University of Cambridge), and was transformed into BL21(DE3) cells as per the instructions of the supplier. Cultures were induced at OD_{600nm} of 0.6 – 0.8 with 1 mM IPTG for ~ 4 h at 37°C. Cells were sonicated in 50 mM Tris-HCl, pH 7.2, and centrifuged at 70,000 x g for 20 min at 4°C. The supernatant was subjected to anion exchange on HiTrap Q FF column and pure

protein eluted by applying a linear NaCl gradient (0 – 0.5 M) in 50 mM Tris-HCl, pH 7.2. Purified *E. coli* E2 lipoyl domain was stored in elution buffer with 5% (v/v) glycerol at -20°C until further use.

3.2.15 Gel shift assay to detect lipoylation

The lipoylation activity of both His-tagged and nontagged LplA and CTD was tested on a simple 15% non-denaturing PAGE gel, which can distinguish between non-lipoylated and lipoylated E2lipD protein due to the difference in their electrophoretic mobilities (Cronan *et al.*, 2005; McManus *et al.*, 2006; Fujiwara *et al.*, 1995). For the assay, all protein components were buffer exchanged (section 3.2.3) into 20 mM HEPES, pH 7.5, 150 mM NaCl. Reaction mixtures contained 100 – 250 µg non-lipoylated E2lipD, 40 mM HEPES (pH 7.5), 10 mM ATP (omitted in negative controls), 10 mM $MgCl_2$, 20 mM D,L-lipoic acid and LplA (10 – 250 µg), CTD (10 – 250 µg) or a mix of both proteins (10 – 250 µg each). The reaction conditions were the same for LplA-CTD Fusion (see chapter 7), adding varying amounts of LplA-CTD Fusion (10 µg – 250 µg) instead of LplA and CTD. Reactions were activated by the addition of ATP and were incubated at room temperature overnight or as stated. Enzyme assays with lipoyl-AMP were set up similarly, replacing D,L-lipoic acid with the alternative substrate and avoiding ATP.

3.2.16 Synthesis of lipoyl-AMP

Lipoyl-AMP was prepared as described by Reed *et al.* (1958) with the modification that AMP.H_2O instead of AMP.$2H_2O$ and DCM in place of acetonitrile were used. A schematic diagram of the synthesis is shown in Fig. 3.7. The synthetic product was analysed by silica thin-layer chromatography (TLC), infrared spectroscopy (IR) and nuclear magnetic resonance (NMR) spectroscopy. The TLC was run in methanol-chloroform (1:1 v/v) on TLC Silica gel 60 on aluminium (Merck Chemicals) and developed using iodine vapours.

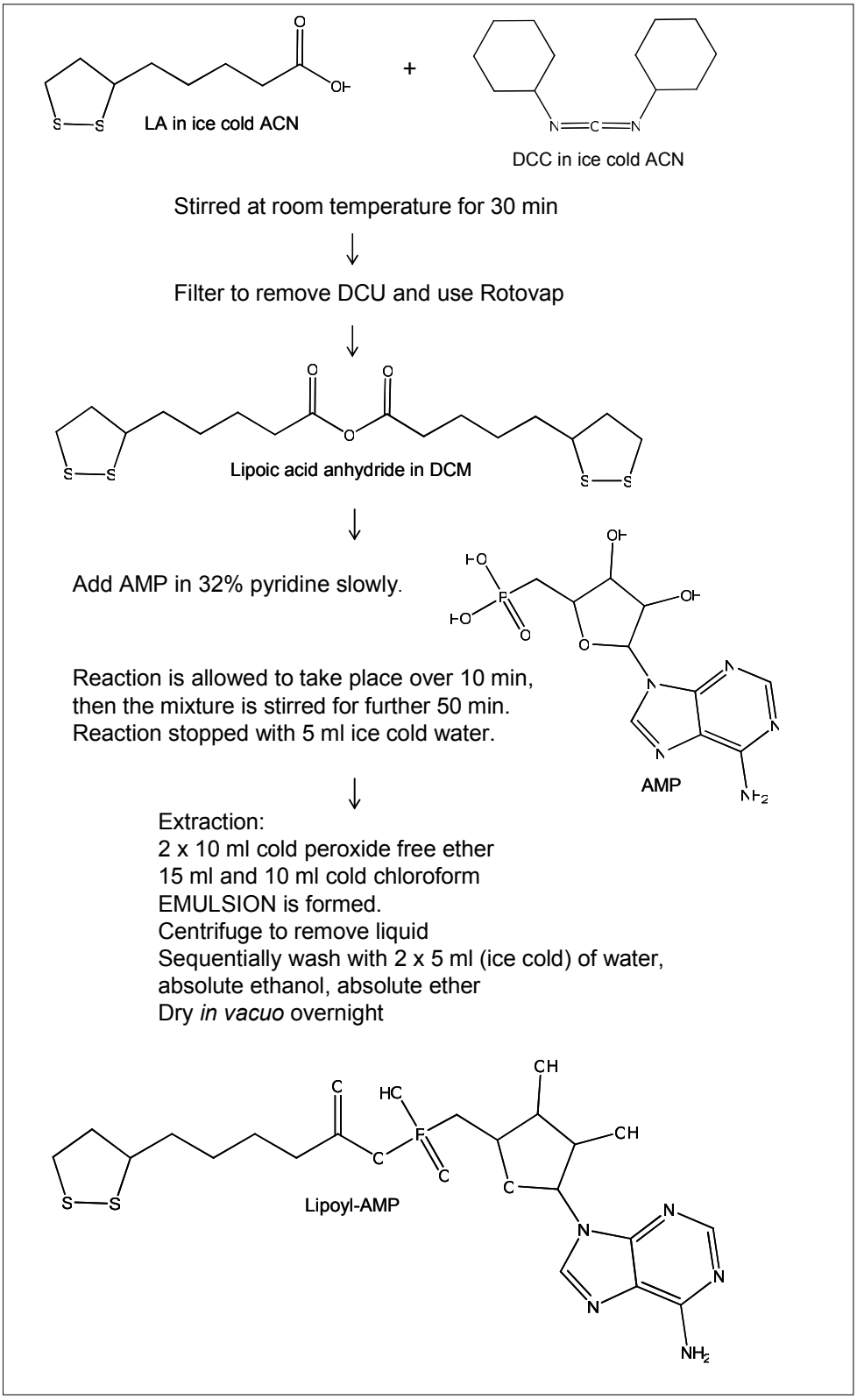


Fig. 3.7 Reaction scheme showing the steps involved in lipoyl-AMP synthesis. LA, lipoic acid; DCC, N,N'-dicyclohexylcarbodiimide; DCU, dicyclohexylurea; ACN, acetonitrile; DCM, dichloromethane.

Infrared spectroscopy was carried out on a Spectrum RX infrared spectrometer (Perkin Elmer). The sample was prepared by mulling (*i.e.* grinding the solid in mineral oil) the reaction product with Nujol^{TM} *(*Trademark of Schering-Plough) and the viscous mixture was applied as a thin-layer between KBr windows.

NMR analysis of lipoyl-AMP was performed on a Varian Mercury VX spectrometer operating at a nominal proton frequency of 400 MHz. For analysis, LA, AMP and lipoyl-AMP were dissolved in D_2O, $CDCl_3$ and DMSO, respectively. ^{1}H NMR spectra were generated for all compounds. ^{31}P NMR spectra were generated for AMP and lipoyl-AMP.

3.2.17 Western blot analysis of lipoylated dihydrolipoamide acyltransferase (E2)

E2 was purified as described in section 3.2.6. 50 µg of whole E2 was incubated with His-LplA (5 µg – 10 µg) and His-CTD (20 µg – 40 µg) (section 3.2.15). The reaction was allowed to proceed overnight and was run on a 15% SDS-PAGE. A wet transfer was done using a Mini Blot system (Biorad) and the transfer buffer (1 L) contained: 100 ml of 10 x transfer buffer (1.92 M glycine, 0.25 M Tris, pH 8.3), 20% (v/v) methanol, 0.0375% (w/v) SDS. The protein was transferred onto an Immobilon-P membrane (Millipore) at 100 V for 1 h at 4°C. Post transfer, the membrane was washed in PBS-T saline buffer (2.68 mM KCl, 136.89 mM NaCl, 10 mM $Na_2HPO_4.2H_2O$, 1.76 mM KH_2PO_4, pH 7.4, 0.1% (v/v) Tween 20) and non-specific binding sites were blocked with 5% (w/v) instant skim-dry milk (Oxoid) in PBS-T for 1 h at room temperature. All incubation steps were done on a gyrating shaker (Janke & Kunkel, Model KS 250, IKA Labortechnik). The membrane was washed thoroughly for 15 min in PBS-T and was incubated in 5% (w/v) milk in PBS-T with a 5000-fold dilution of anti-lipoic acid antibody (Calbiochem) for 1 h at room temperature. The membrane was washed as before and incubated with anti-rabbit IgG (whole molecule)-peroxidase antibody developed in goat (Sigma), 5000-fold dilution in 5% (w/v) milk in PBS-T for 1 h at room temperature. After the final incubation, the membrane was washed in PBS-T for 20 min and incubated with Immobilon Western Chemiluminescent HRP Substrate (Millipore) for 5 min. The Western blot was developed using X-ray film (Fujitsu) and a developer.

3.2.18 Mass Spectrometry

Mass spectrometry (MS) analysis was done at the University of Bath using a micrOTOF^{TM} electrospray time-of-flight (ESI-TOF) mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany), which can be used to measure accurate mass to 5 ppm externally calibrated. Lipoylation reactions of E2lipD (1 mg) with LplA:CTD in a 1:1 molar ratio were carried out overnight; negative controls contained no ATP. The reaction mixtures were then dialysed against 20 mM Tris-HCl, pH 8.5, and 50µl of 50% (v/v) acetonitrile-0.1% formic acid were added to 100µl of the protein sample prior to analysis. Samples were analysed by direct infusion at 4 µl/min from a syringe pump. The electrospray source was used in positive ion mode, with a dry-gas flow-rate of 4 L/min at 20°C. The nebuliser pressure was set at 0.4 bar, with a capillary voltage of 4500V, a capillary exit of 180V and a skimmer 1 voltage of 90V. The ESI-TOF scanned a mass range from 250 – 3000 m/z and the average mass spectrum was externally calibrated against 5 mM sodium formate clusters using the data processing package DataAnalysis^{TM} (Bruker Daltonik GmbH, Bremen, Germany).

3.2.19 Protein concentration estimation

Protein concentrations were determined by measuring A_{280nm} of protein solutions in 6 M guandine-HCl (Kelly *et al.*, 2005). Extinction coefficients were calculated from the protein sequence using the ExPASY ProtParam tool (Swiss Institute of Bioinformatics).

3.3 Biophysical methods

3.3.1 Pull-down assays

The pull-down assay is an *in vitro* method that is based on affinity chromatography. It allows confirmation and detection of previously unknown interactions between two or more proteins and is complementary to techniques such as yeast two-hybrid, enzyme-linked immunosorbant assay (ELISA) and analytical ultracentrifugation. If the nature of the protein-protein interaction is known, this technique allows purification of a target protein from an unfractionated cell extract. For a pull-down assay, sufficient amounts of pure tagged protein, the bait,

are required to capture and "pull-down" the binding partner (the prey). Preferably the prey is non-tagged or has a different tag from the bait protein. The principle of pull-down assays has been widely employed to identify and/or confirm protein-protein interactions and to complement protein biophysical analyses (Promega Protein Interaction Guide (2007), chapter 3, [Online]: Available at http://www.promega.com/guides/protein.interactions_guide/protein.interactions_gde.pdf (accessed December 7th, 2008).

Any formation of a complex between LplA, CTD and E2lipD was analysed by employing a pull-down assay. His-LplA (the bait) was loaded onto a HisTrap FF column in buffer A (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) and the column was washed with buffer to remove any unbound His-LplA. ntCTD (the prey) in 20 mM Tris-HCl, pH 7.5, was also passed over the same column in buffer A followed by extensive washing with buffer to remove any unbound ntCTD. Protein(s) were then eluted with a linear gradient of imidazole (0 – 0.5 M) in 20 mM Tris-HCl, pH 7.5, 0.3 M NaCl, and fractions were analyzed on 15% (w/v) SDS-PAGE. The interaction of His-tagged E2LipD with ntLplA or ntCTD was also studied in a similar way. The activity of the purified His-LplA-ntCTD complex was tested in gel shift assays as described in section 3.2.15. Purified complexes were also analyzed using dynamic light scattering (section 3.3.3).

3.3.2 Surface plasmon resonance (SPR)

All SPR experiments were performed on Biacore® T100 or Biacore® X-100 (GE Healthcare) at 25°C in the analyte buffer. For immobilization of ligands, the sensor chip surface was activated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and N-hydroxysuccinimide (NHS), which had been premixed in a 1:1 ratio of equal concentration. The surface was blocked with ethanolamine three times as per the manufacturer's instructions. The HBS-EP buffer (GE Healthcare) contained 0.01 M HEPES, pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20.

The CM5 chip has a ~ 100 nm thick surface layer made up of carboxymethylated dextran. This dextran matrix provides a high surface capacity for immobi-

lizing many different ligands (Biacore® sensor surface Handbook, edition October 2003).

Ligand penta-His antibody (50 µg/ml) was immobilized onto a CM5 chip in 10 mM sodium acetate buffer, pH 5.0 yielding RU = 600. The surface was charged with Ni^{2+} for 60 s before the analyte, His-tagged LplA (100 µg/ml) in HBS-EP buffer, was injected for 120 s at a flow rate of 5 µl/min. ntCTD in the presence and absence of 5 mM $MgCl_2$ in HBS-EP buffer was used as a secondary analyte and passed over the already bound His-LplA. Similarly, His-tagged E2lipD was bound to the penta-His antibody on the CM5 chip. As analyte, ntLplA (100 µg/ml) by itself or in the presence of ntCTD (100 µg/ml) was passed over His-E2lipD ligand.

Ligand ntLplA (50 µg/ml) was immobilized onto a CM5 chip in 10 mM sodium acetate buffer, pH 5.0 yielding RU = 567.1. Analyte was ntCTD (4 mg/ml) in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$. A serial dilution of ntCTD was prepared from the stock solution (4 mg/ml). ntCTD was injected for 60 s (association phase) and dissociation phase was monitored for 600 s at a flow rate of 30 µl/min. Regeneration of the sensor chip surface was achieved by using 10 mM NaOH for 30 s.

Ligand ntCTD (50 µg/ml) was immobilized onto a CM5 chip in 10 mM sodium acetate buffer, pH 5.0 yielding RU = 400. Analytes were ntLplA (3 mg/ml) and His-E2lipD (4 mg/ml). A serial dilution of ntLplA was prepared in HBS-EP buffer, pH 7.4, 5 mM $MgCl_2$, NSB (non-specific binding reducer; GE Healthcare). The maximum and minimum concentrations of ntLplA used were 10 mg/ml and 0.04 mg /ml, respectively. His-E2lipD was diluted to 0.2 mg/ml in HBS-EP buffer, pH 7.4, 5 mM $MgCl_2$, and used as the analyte. In addition, His-E2lipD (0.2 mg/ml) and ntLplA (3 mg/ml) were mixed in a ratio of 1:1 and used as analyte. The flow rate was 30 µl/min. Washing with running buffer regenerated the sensor chip surface.

3.3.3 Dynamic light scattering (DLS)

All DLS measurements were performed using a Zetasizer Nano S from Malvern Instruments Ltd. (Malvern, UK) using a low volume precision quartz cuvette (105.251.005-QS, Hellma UK Ltd.). Prior to DLS measurements, protein solutions (≥ 1 mg/ml) in 20 mM Tris-HCl, pH 8.0 and 0.2 M NaCl solution were filtered through a 0.02 µm membrane filter (Whatman® Anotop 10, Fisher Scientific) or centrifuged at 13,000 x g for 20 min at 4°C to remove dust particles. DLS measurements were carried out at 25°C or 55°C in triplicate. Melting points were determined over a temperature range of 20°C to 70°C.

Analysis of purified OADHC components: All purified protein samples (≥ 1 mg/ml) were in 20 mM sodium phosphate, pH 7.4. Purified E1 and E2 were filtered through 0.02 µm membrane filters. Purified E3 was filtered through a 0.22 µm membrane filter (Millipore®) as E3 filtered through 0.02 µm filters was unsuitable for measurement. Each measurement was done in triplicate. Two operating temperatures were used, 25°C or 55°C. For measurements at 25°C, volumes of 12 – 20 µl were sufficient, whereas for measurements at 55°C at least 100 µl of sample was required to counter the effects of evaporation. The complex was assembled in the E1 : E2 : E3 molar ratio of 3 : 1 : 0.1, which has been reported to give optimum enzyme activity (Heath *et al.*, 2007). The concentration of assembled complex in DLS was 1 mg/ml. To measure the effect of NaCl on protein oligomerization, samples were heated at 55°C for 10 min in the presence or absence of NaCl (0 – 350 mM). Samples heated without NaCl were supplemented with NaCl after they had cooled down to 25°C. Samples were then analysed by DLS at 25°C and 55°C.

Analysis of purified LplA, CTD and LplA-CTD complex: All purified recombinant proteins were buffer exchanged into 20 mM Tris-HCl, pH 8.0, 0.2 M NaCl and centrifuged at 13,000 x g for 30 min at 20°C. The protein concentration of purified His-LplA, His-CTD and ntCTD was approximately 1 mg/ml. The complex of His-LplA and His-CTD was obtained from gel permeation chromatography (section 3.3.5).

M_r values of proteins were derived from the measured hydrodynamic radii using the Protein Utilities feature of the DISPERSION TECHNOLOGY software, version 4.10, supplied with the instrument.

3.3.4 Analytical ultracentrifugation (AUC)

All analytical ultracentrifugation experiments were carried out on a Beckman XL-A analytical ultracentrifuge (Beckman-Coulter, CA). Sedimentation velocity experiments were carried out at 15 000 r.p.m., and cells were scanned every 5 min at 280 nm. For sedimentation velocity analysis of the BCOADHC, the buffer was 20 mM sodium phosphate, pH 7.5, 10% (v/v) glycerol and 100 mM NaCl. The set temperature on the centrifuge was 40°C, and the solution densities were directly measured at this temperature using an Anton-Paar DMA 5000 high-precision density-meter. Sedimentation velocity distributions were obtained by the c(s) method (Schuck, 2000) using the program Sedfit. Data were then directly fitted using the finite element solution of the Lamm equation in Sedfit (Schuck, 1998) to give values of the sedimentation coefficient and of M_r.

3.3.5 Gel permeation chromatography (GPC)

GPC experiments were performed using a Superdex 200 10/300 GL Tricorn column (Amersham Pharmacia). The column was equilibrated in the running buffer 20 mM Tris-HCl, pH 7.5, 150 mM NaCl or running buffer without NaCl. The column was calibrated with protein standards (Sigma) in the corresponding running buffer. Prior to loading on the column, proteins were dialyzed into the corresponding running buffer with or without NaCl (section 3.2.3). Proteins were eluted with a flow rate of 1 ml/min.

3.3.6 Isothermal titration calorimetry (ITC)

For ITC, His-CTD and His-LplA were extensively dialysed into 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$ (section 3.2.3) and the final dialysis buffer was saved for later use. Either His-CTD or His-LplA was used in the bomb with the other protein as ligand. Experiments were run at either 25°C or 40°C. All experiments were carried out using a MicroCal VP-ITC calorimeter (MicroCal, Northampton, USA).

3.3.7 Nuclear magnetic resonance (NMR) spectroscopy

His-tagged ^{15}N-labelled CTD was produced by expression in M9 minimal medium (Table 2) supplemented with 1g/L $^{15}NH_4Cl$ as the sole nitrogen source. The rest of the protocol was the same as described in previous sections for different proteins.

For NMR analysis, purified protein samples were buffer exchanged into different buffers (see 3.2.3). NMR data were acquired at 37°C or 55°C on a Varian Unity INOVA spectrometer operating at a nominal proton frequency of 600 MHz, using a triple resonance 5 mm probe equipped with z-axis pulsed field gradients. $^{1}H–^{15}N$ heteronuclear single quantum coherence (HSQC) spectra were generated and processed using Sparky NMR Assignment and Integration Software (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Table 2 M9 minimal medium composition

M9 minium medium contains (1L)		**Trace element solution in distilled H_2O, adjusted to pH 8.0 (1 L)**	
Na_2HPO_4	6.74 g	$CaCl_2.2H_2O$	5.5 g
KH_2PO4	3 g	$MnSO_4.H_2O$	1.4 g
NaCl	0.05 g	$CuSO_4.5H_2O$	0.4 g
$^{15}NH_4Cl$	1.0 g	$ZnSO_4.7H_2O$	2.2 g
$MgSO_4$ (1 M)	1.0 ml	$CoSO_4.7H_2O$	0.45 g
$CaCl_2$ (1 M)	0.1 ml	$Na_2MoO_4.2H_2O$	0.26 g
Trace element solution	0.65 ml	H_3BO_4	0.4 g
Biotin	0.1% w/v	KI	0.26 g
Thiamine	0.1% w/v	EDTA	5 g
D-glucose	20% w/v	$FeSO_4.7H_2O$	3.75 g

3.3.8 Crystal screens

The His-LplA-ntCTD complex purified from the pull-down assay was used for crystallization. Two different methods were used:

Hanging drop method: Three different conditions were tested per well by mixing 1 µl of protein solution with 1 or 2 µl of reservoir buffer or just 1 µl of protein solution by itself on one glass cover slip. The coverslip was then inverted and sealed over the reservoir well using vacuum grease, and the plate was stored at 18°C. The different conditions tested for co-crystallization were 50, 55, 60, 65,

70, and 75% tacsimate solutions at pH 7.0 and pH 8.0, Hampton CryoCrystal-screen Nr. 6 and Nr. 40, and Molecular dimensions crystal screen I, Nr. 17 and Nr. 41.

Sitting drop method: Crystal screens were set up using the Phoenix Protein crystallography robot (Art Robbin's instruments). The whole crystal screen 2, except condition 49 and 50, was used and the plate was stored at 16°C.

CHAPTER 4

Gene and protein sequence analysis of *Thermoplasma acidophilum* lipoate protein ligase and its putative C-terminal domain

4.1 Introduction

The characterization of the recombinant BCOADHC from *Tp. acidophilum* showed that specific growth conditions are required for the *E. coli* host to carry out lipoylation of the E2 component (Heath *et al.*, 2007). As lipoylation is a crucial pre-requisite for the activity of the complex, attention was drawn to the innate *Tp. acidophilum* lipoylation machinery. Previous genome analyses showed that *Tp. acidophilum* has one putative lipoate protein ligase (LplA) gene and the crystal structure of the encoded *Tp.* LplA shared similarities with other biotinylating/lipoylating enzymes (McManus *et al.*, 2006). However, *in vitro* activity of recombinant *Tp.* LplA could not be fully established. The presence of the reaction intermediate lipoyl-AMP as part of the crystal structure indicated at least partial enzyme activity (Kim *et al.*, 2005). Contradictory to this, *in vitro* lipoylation assays failed to detect enzyme activity and it was suggested that an additional protein/enzyme component might be required for full activity of the *Tp.* LplA (McManus *et al.*, 2006). This chapter describes the characterisation of gene and protein sequences of the putative *Tp.* LplA and discusses the potential of exploring a second gene to reconstitute a fully functional lipoate protein ligase.

4.2. Results

4.2.1 Gene sequence analysis of the *Thermoplasma* LplA and its putative C-terminal domain

Protein sequences of the *E. coli* LipA, LipB and LplA and the *Thermoplasma* BCOADHC components were used in BLASTP searches against archaeal genomes (Table 4.1). The *E. coli* protein sequences identified the open reading frame (ORF), *ta0514*, which codes for the putative *Tp.* LplA. In an attempt to refine the search, the protein sequences of lipoate protein ligases with known

crystal structures were included in BLASTP searches. Using the *Streptococcus pneumoniae* LplA (PDB ID: 1vqz) protein sequence, a second ORF, *ta0513m*, was identified exactly upstream of *ta0514*. On the protein level, the *Tp.* LplA was found to align with the N-terminal domain of the *S. pneumoniae* lipoate protein ligase, whereas the predicted protein product of *ta0513m* aligned with the C-terminal domain of the *S. pneumoniae* lipoate protein ligase (Fig. 4.1). Notably, the two ORFs, *ta0514* and *ta0513m*, overlap with each other, as the stop codon of the *ta0513m* (TGA) shares the adenine with the start codon of the *ta0514* (ATG). From these sequence alignments it appears that both gene products might reconstitute a functional *Tp.* LplA. The two ORFs were analysed for transcription/translation motifs to gain any information that would indicate any association. Based on the sequence alignment with the *Sp. pneumoniae* LplA (Fig. 4.1), *ta0514* will be referred to as *lpla* and *ta0513m* as *ctd* and the gene products as *Tp.* LplA and CTD respectively.

Archaea use characteristic transcription initiation signals that help to guide the RNA polymerase (Werner, 2007). The presence of these features in an assigned ORF strengthens the hypothesis for its *in vivo* use. In archaeal promoters, two major sequence elements have been identified. The TATA-box, an A/T-rich sequence approximately 25 base pairs upstream of the transcription start site, is recognized by the TATA-binding protein (TBP), which in turn is required for RNA polymerase recruitment. Secondly, the B-recognition element (BRE), which is located adjacent and upstream of the TATA box, is recognized by transcription factor B (TFB) and determines promoter strength and direction of transcription (Bartlett, 2005; Gregor and Pfeifer, 2005; Qureshi and Jackson, 1998; Kosa *et al.*, 1997; Littlefield *et al.*, 1999). The archaeal TATA-like box sequence is TTTA[A/T]. The TATA box combined with an upstream BRE has the consensus of 5`cRnaAntTTTAWAtr3` where R can be any purine, W any pyrimidine, n any base, and lower case letter is the most common nucleotide in a variable position (Littlefield *et al.*, 1999; Qureshi and Jackson, 1998). Both *lpla* and *ctd* comprise a potential TATA-like A box with a BRE adjacent upstream of it (Fig. 4.2). This indicates that the DNA is present in a form accessible for the translation machinery. Interestingly, the elements for the *lpla* gene fall within the *ctd* gene.

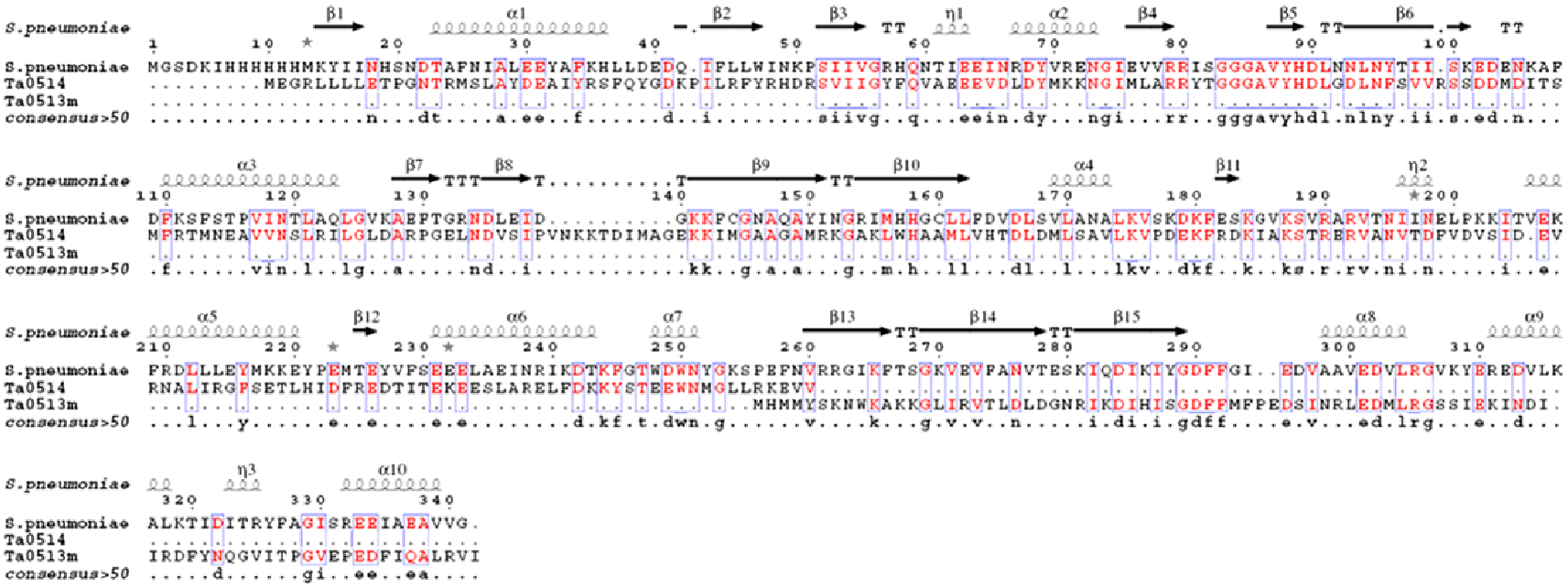


Fig. 4.1 Sequence and structural alignment of *S. pneumonia* LplA with the putative *Tp. acidophilum* LplA (Ta0514) and CTD (Ta0513m). The sequence alignment was done using MultiAlin (Corpet, 1988). Secondary structure assignment of the *S. pneumoniae* LplA was done in ESPript (Gouet *et al.*, 1999).

			E1 alpha	E1 beta	E2	E3	LipA	LipB	Tp. LplA	CTD
EURYARCHAEOTA	**Thermoplasmales**	*Thermoplasma acidophilum*	Ta1438	Ta1437	Ta1436	Ta1435	x	x	Ta0514	Ta0513
		Thermoplasma volcanium	TVN0102	TVN0101	TVN0100	TVN0099	x	x	TVN0991	TVN0990
		Picrophilus torridus	PTO0549	PTO0548	PTO0547	PTO0546	PTO0550	PTO0545	x	x
	Halobacteriales	*Halobacterium sp. NRC-1*	VNG1926G	VNG2218G	VNG2219G	VNG2220G	VNG2216G	x	VNG6256G	VNG6255C
		Haloarcula marismortui	rrnAC2957	rrnAC2956	rrnAC2955	rrnAC2953	rrnAC2959	x	rrnAC2019	x
CRENARCHAEOTA	**Thermoprotales**	*Pyrobaculum aerophilum*	PAE2644	PAE2646	PAE2648	PAE2649	PAE2643	PAE2651	x	x
		Pyrobaculum calidifontis	Pcal_1405	Pcal_1404	Pcal_1403	Pcal_1402	Pcal_1406	Pcal_1401	x	x
		Pyrobaculum arsenatium	Pars_1189	Pars_1188	Pars_1187	Pars_1205 Pars_1572	Pars_1184	Pars_1183	x	x

Table 4.1. Archaeal genomes containing genes for putative OADHC components, LplA, LipA and LipB. A BLASTP search using the protein sequences of *Tp. acidophilum* E1α (Ta1438), E1β (Ta1437), E2 (Ta1436) and E3 (Ta1436) was done against the 48 archaeal genomes accessible in Pubmed. To identify putative lipoylation enzymes in archaeal genomes, BLASTP searches were done using the protein sequences of *E. coli* LipA (Uniprot accession (P60716), *E. coli* LipB (Uniprot accession P60720) and *Tp. acidophilum* LplA (Ta0514) and CTD (TA0513m). The table shows archaeal genomes in which all OADHC components are present in clusters/ putative operons as well as potential lipoylation genes. Separate ORFs for LplA and its putative C-terminal domain was analyzed using the identified CTD (*ta0513m*) from *Tp. acidophlium*.

Fig. 4.2 Gene sequences of *Tp. acidophilum* LplA (*ta0514*; *lpla*) and its putative C-terminal domain (*ta0513*; *ctd*). The gene sequence data are shown as present in the genome, including some upstream and downstream sequence data, surrounding the two ORFs. The first base of the start codons of the *ctd* and *lpla* gene are indicated by a circle and star, respectively. The stop codon of *ctd* (TGA) and start codon of lpla (ATG) are in capital letters with the shared base (A) in bold. The *lpla* stop codon (TAG) is indicated by a cross. The TATA-like A box is shown in bold lower-case letters. The Shine-Dalgarno sequence is shown in bold capital italics. BRE elements are underlined. Proposed poly-T stop signals underlined by a double line. Arrows below the sequence indicate stems of putative mRNA secondary structure.

Like translation initiation signals, the presence of transcription signals is often used as an indicator for a functional gene. An element that aids translation initiation is a purine rich motif, the Shine-Dalgarno (SD) sequence (Kozak, 2005). In *Tp. acidophilum* the conserved core is represented as GGAGG, but variations including GTGA, GGTG, AGGT and GAGG are common and do not compromise the effectiveness of SD sequences as long as the base-pairing potential with the anti-SD sequence and its spacing from the start codon are efficient (Ma *et al.*, 2002). Actually, each motif can bind to the underlined part of the 16S rRNA sequence AUCACCUCC (Torarinsson *et al.*, 2005). The *ctd* gene has no

SD sequence, whereas the *lpla* gene has a slight variation, GAGG, 12 base pairs upstream of the start codon. Although the presence of a SD sequence is beneficial for gene expression levels, genes without an SD sequence are common. *Tp. acidophilum* has been found to have many leaderless transcripts and its genome is known to be low in SD motifs (Chang *et al.*, 2006; Torarinsson *et al.*, 2005; Karlin *et al.*, 2005; Ma *et al.*, 2002).

In addition to translation motifs, the secondary structure of mRNA transcripts has been found to affect translation initiation (Kozak, 2005; Hall *et al.*, 1982). As pointed out above, the *ctd* and *lpla* genes are connected via a 1 base pair overlap positioned in the *ctd* stop and *lpla* start codon. Hence, if both genes were transcribed together, correct translation of the *lpla* gene would require a -1 base pair frameshift. Secondary mRNA structures have been shown to aid frameshifting by stalling the RNA polymerase (Giedroc and Cornish, 2008; Kozak, 2005). To assess any potential involvement of mRNA structures on *ctd-lpla* translation, the last 127 base pairs of the *ctd* 3` end followed by the first 139 base pairs of the *lpla* 5` end were analysed for the presence of mRNA secondary structures. This analysis indicated a possible stem-loop structure exactly downstream of the *lpla* SD sequence, with the overlap region between the *ctd* stop and *lpla* start codon forming part of the stem (Fig. 4.2 and Fig. 4.3). The presence of such a structure in the stop-start-codon region has potential to cause stalling of the RNA polymerase at the *ctd* stop codon allowing reinitiation at the *lpla* start codon.

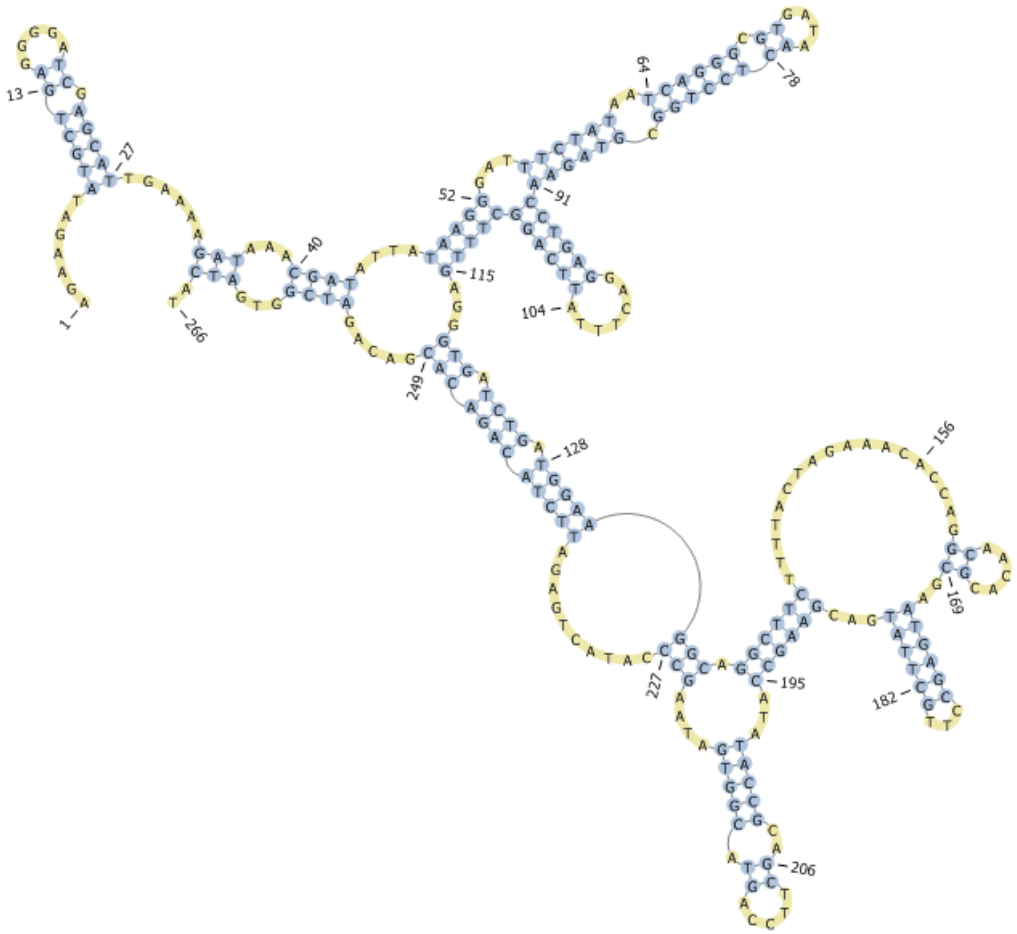


Fig. 4.3 Two-dimensional visualization of predicted secondary mRNA structures of a *ctd-lpla* transcript (Webservice pknotsRG, BiBiServ Bielefeld University bioinformatics Server, for description of algorithms see Reeder and Giegerich, 2004). The submitted sequences consisted of the last 127 base pairs of the *ctd* 3`end followed by the first 139 base pairs of the *lpla* 5` end. The second base pair of the *lpla* is indicated as residue 128. The diagram illustrates the 12 base pair stem structure encompassing the stop codon of *ctd* and start codon of *lpla*. The adenine shared by the *ctd* stop codon and *lpla* start codon is not paired with another residue in the stem structure.

4.2.2 Protein sequence analysis of the *Thermoplasma* LplA (*Tp.* LplA) and its putative C-terminal domain (CTD)

The predicted protein sequences of the *lpla* and *ctd* gene were analyzed for functional motifs and compared with LplAs whose structures have been deposited in the Protein Data Bank (PDB). The *Tp.* LplA protein sequence of 262 amino acids is 22.3 – 24.5% shorter than other lipoate protein ligases whose crystal structures have been solved. Despite differences in length and the possibility of shorter LplAs lacking functional domain parts, shorter LplAs have been found to be as active as their longer counterparts. One example of a functional

LplA with only 270 amino acid residues is the LplA from Oryza Rice (Kang *et al.*, 2007).

A Pfam search of the protein sequence of *Tp.* LplA revealed a BPL_LipA_LipB domain (Pfam-A motif PF03099, amino acids 47 – 169) (Fig. 4.4) which is in agreement with other LplAs (Reche, 2000). A putative cAMP/cGMP-dependent protein kinase phosphorylation site motif is present in *Tp.* LplA (Fig. 4.4), which is thought to be directly involved in the formation of lipoyl-AMP intermediates from lipoic acid and ATP (Kang *et al.*, 2007).

```
1    MEGRLLLLET PGNTRMSLAY DEAIYRSFQY GDKPILRFYR HDRSVIIGYF QVAEEEVDLD
61   YMKKNGIMLA RRYTGGGAVY HDLGDLNFSV VRSSDDMDIT SMFRTMNEAV VNSLRILGLD
121  ARPGELNDVS IPVNKKTDIM AGEKKIMGAA GAMRKGAKLW HAAMLVHTDL DMLSAVLKVP
181  DEKFRDKIAK STRERVANVT DFVDVSIDEV RNALIRGFSE TLHIDFREDT ITEKEESLAR
241  ELFDKKYSTE EWNMGLLRKE VV
```

Fig. 4.4 Pfam analysis of *Tp.* LplA encoded by gene locus *ta0514* (NP_393990.1; 262 amino acids). The residue number is given on the left. The BPL_LipA_LipB domain is underlined. The putative cAMP dependent protein kinase phosphorylation site (RRYT) is shown in bold.

In addition to the correlation of the BPL_LipA_LipB, a structural comparison of the crystal structures showed that the *Tp. LplA* has the same overall fold and aligns with the N-terminal domain of the *S. pneumoniae* LplA (Fig. 4.5.). This comparison also highlights the lack of an individually folded C-terminal domain in the *Tp.* LplA.

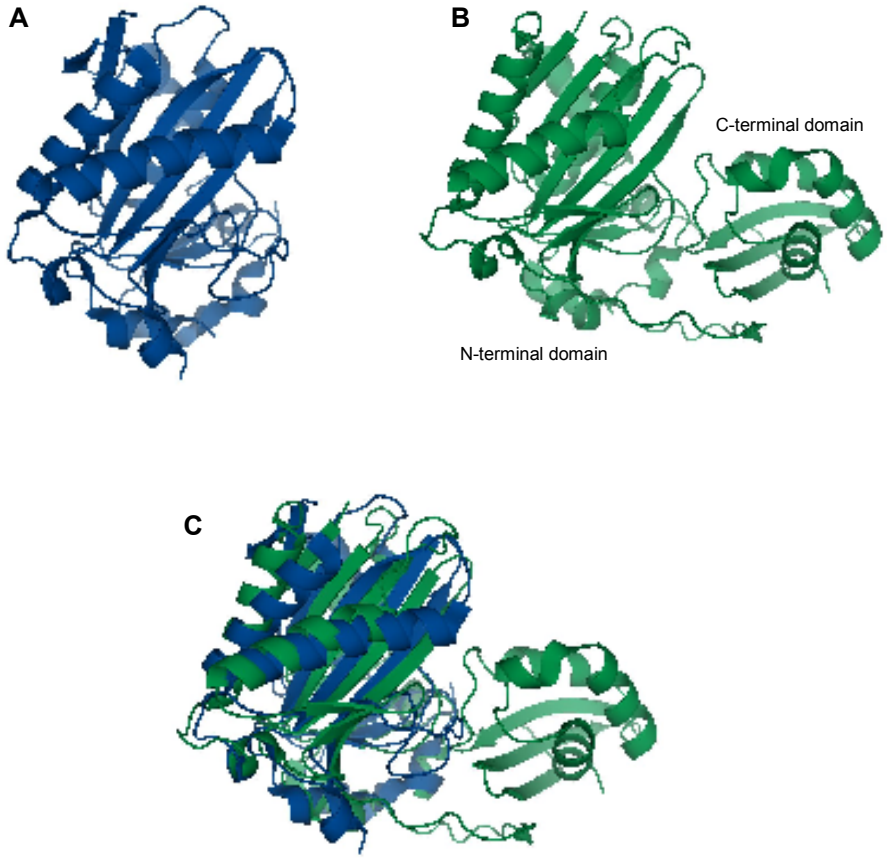


Fig. 4.5 Structural comparison of the domain structure of lipoyl-protein ligases. **A**: The crystal structure of *Tp. acidophilum* LplA (PDB accession: 2ars). **B**: The crystal structure of *Streptococcus pneumoniae* LplA showing the N- and C-terminal domains (PDB accession: 1vqz, Joint Center for Structural Genomics, to be published). **C**: Overlay of A and B. Structures were generated using PyMOL (DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org).

Since the start of this project, the *Tp. acidophilum* genome has been reannotated resulting in two ORFs annotated as *ta0513m* and *ta0513* (Fig. 4.6). The ORF annotated as *ta0513* codes for 94 amino acids, whereas *ta0513m* corresponds to only 88 amino acids. Moreover, two more start codons just downstream of *ta0513m* could also represent the ORF. Based on the position of transcription initiation elements and the sequence alignments in which the CTD barely overlaps with the *Tp.* LplA but matches the *S. pneumoniae* LplA C-terminal, it was concluded that the annotated *ta0513m* was the most likely candidate to be used *in vivo* and was cloned and expressed recombinantly for activity studies.

```
I                    II      III IV
atggttctcaattatactatgcatatgatgtacagcaagaactgg
M  V  L  N  Y  T  M  H  M  M  Y  S  K  N  W

aaggcgaagaagggactcataagggtaacgctggatctggacggg
K  A  K  K  G  L  I  R  V  T  L  D  L  D  G

aatagaataaaggacatacacatttccggagatttcttcatgttt
N  R  I  K  D  I  H  I  S  G  D  F  F  M  F

ccggaagattccataaacagactagaagatatgctgaggggatcg
P  E  D  S  I  N  R  L  E  D  M  L  R  G  S

agcattgaaaagataaacgatattataagggatttctataatcag
S  I  E  K  I  N  D  I  I  R  D  F  Y  N  Q

ggcgtgataactcctggcgtagaacctgaggactttattcaggct
G  V  I  T  P  G  V  E  P  E  D  F  I  Q  A

ttgagggtgatctga
L  R  V  I  *
```

Fig. 4.6 Potential ORFs coding for CTD and the corresponding protein sequence. Four possible start codons (ATG) of the *ctd* gene are shown in bold and numbered with Roman numerals; the stop codon (TGA) is in bold and indicated by a star. The reannotated ORF *ta0513*, coding for 94 amino acids, starts at I. The ORF *ta0513m*, coding for 88 amino acids, starts at II and was expressed recombinantly for analysis.

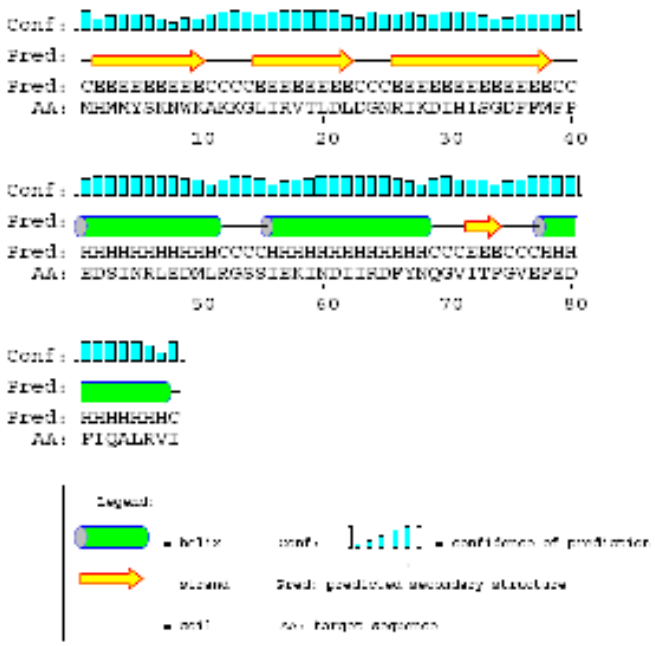


Fig. 4.7 Secondary structure prediction of CTD using PSIPRED (Jones, 1999).

A Pfam search of the putative C-terminal domain of the *Thermoplasma* LplA identified the presence of a Lip_prot_lig_C domain (Pfam-A motif PF10437, amino acids 1 to 85). This motif is also referred to as bacterial lipoate protein ligase C-terminus and is associated with the BPL_LipA_LipB motif (see Fig.

4.4). This domain is thought to be stable and it has been hypothesised that it may recognise the lipoyl domain and/or transfer the lipoyl group from the intermediate (McManus *et al.*, 2006). *S. pneumoniae* LplA also has this domain at its C-terminus, which may support a role of CTD as the C-terminal domain of the *Tp.* LplA in lipoylation. In a sequence alignment (Fig 4.1) the CTD matched well with the C-terminal of *S. pneumoniae* LplA. Assignment of the secondary structure elements to the CTD sequence using PSIPRED highlights the same position of secondary structure elements as found in the *S. pneumoniae* C-terminal domains with solid confidence (Fig. 4.7).

4.2.3 Attempts to solve the protein structure of CTD using NMR spectroscopy and X-ray crystallography

The protein sequence analyses of CTD suggested that it might adopt a fold similar to the C-terminal domain of other LplAs. However, the reality may be different as we are proposing to study CTD as an individual protein. Given the small size of CTD, NMR spectroscopy was deemed fit as a technique to try to investigate its structure. In order to achieve sufficient signal to noise ratio for protein structure determination by NMR, a concentrated (approximately 1 mM) ^{15}N labelled protein solution is required. ^{15}N labelled protein can be recombinantly produced by growing the cells in M9 minimal medium (section 3.3.7) thereby compromising the yield of protein in comparison to expression in standard LB medium. In the case of CTD, the His-tagged version gave a higher yield than its non-tagged counterpart. NMR data on ^{15}N His-CTD in 20 mM MES, pH 6.5, 100 mM NaCl were acquired at 37°C (Fig. 4.8, A). Acquiring the $^{1}H–^{15}N$ heteronuclear single quantum coherence (HSQC) spectrum took 3.25 h and was not suitable for any structural assignments. The acquisition time was increased to 17 h to see if this gives a stronger overall signal, but it did not improve the resolution of the spectrum (Fig. 4.8, B). The main problem at this stage was a low protein concentration, which did not give a good signal to noise ratio in 30 min of acquisition. Further, the peak overlap in the central part of the spectrum and the variation in peak intensity might have been a result of protein aggregation, conformational heterogeneity and/or impurities.

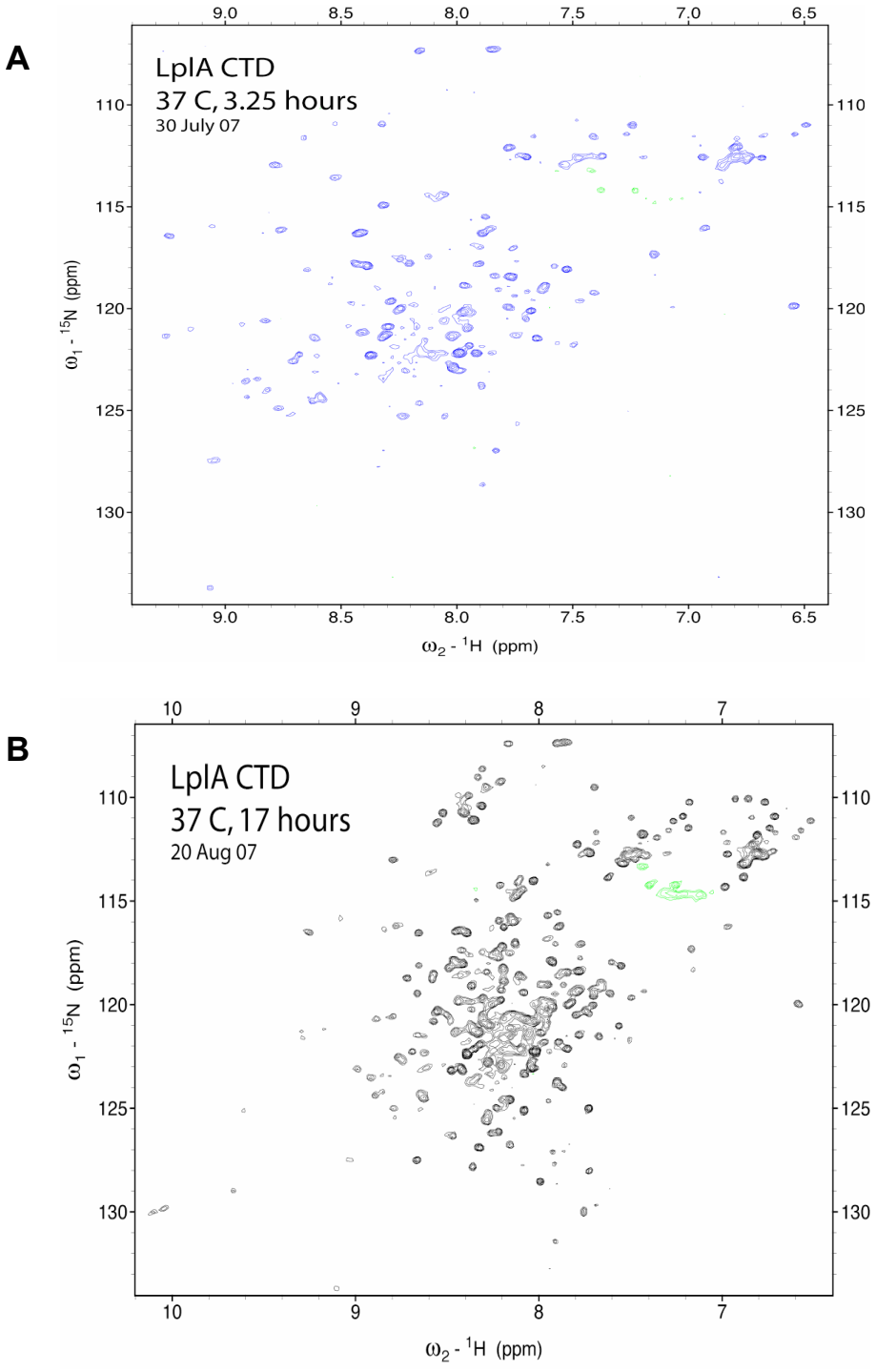


Fig. 4.8 HSQC spectra of His-tagged CTD in 20 mM MES, pH 6.5, 100 mM NaCl at 37°C. **A**: Acquisition time 3.25 h. Number of scans: 48, proton spectral width: 8000 Hz, nitrogen spectral width: 1763 Hz. **B**: Acquisition time 17 h. Number of scans: 256, proton spectral width: 8000 Hz, nitrogen spectral width: 1763 Hz.

The initial HSQC spectrum indicated that there was a population of folded molecules prone to aggregation and/or precipitation (Fig. 4.8). As CTD is a thermophilic protein, the NMR experiment was repeated at 50°C to see the effect of temperature on protein stability and conformation (Fig. 4.9). The change in temperature improved the signal dispersion in the HSQC spectrum but the degradation prevailed as the peak number and peak intensity still varied.

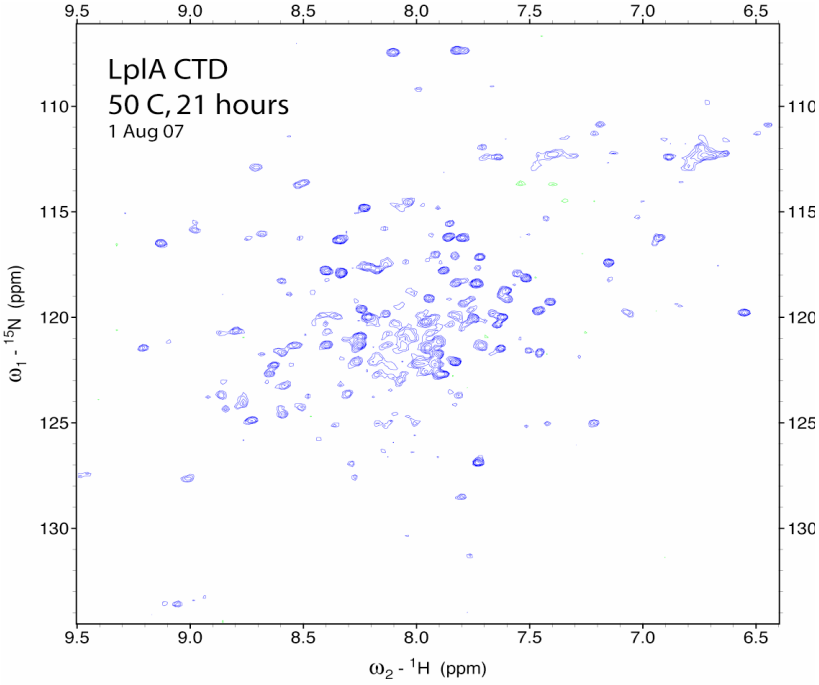


Fig. 4.9 HSQC spectrum of His-tagged CTD in 20 mM MES, pH 6.5, 100 mM NaCl at 50°C. Number of scans: 320, proton spectral width: 8000 Hz, nitrogen spectral width: 1763 Hz.

3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) was used as an additive to try to prevent aggregation (Anglister *et al*., 1993). Its effect on His-CTD HSQC spectra was tested at 37°C at pH 6.5 (Fig. 4.10, A). The effect of different pH conditions on the HSQC of His-CTD was also tested. His-CTD has a theoretical pI of 6.75. Nonetheless, any pH value lower than pH 6.5 caused rapid degradation and aggregation. As His-CTD is stable during purification at pH 8.0, this pH value was chosen to acquire the HSQC spectrum in the presence of 20 mM CHAPS (Fig. 4.10, B).

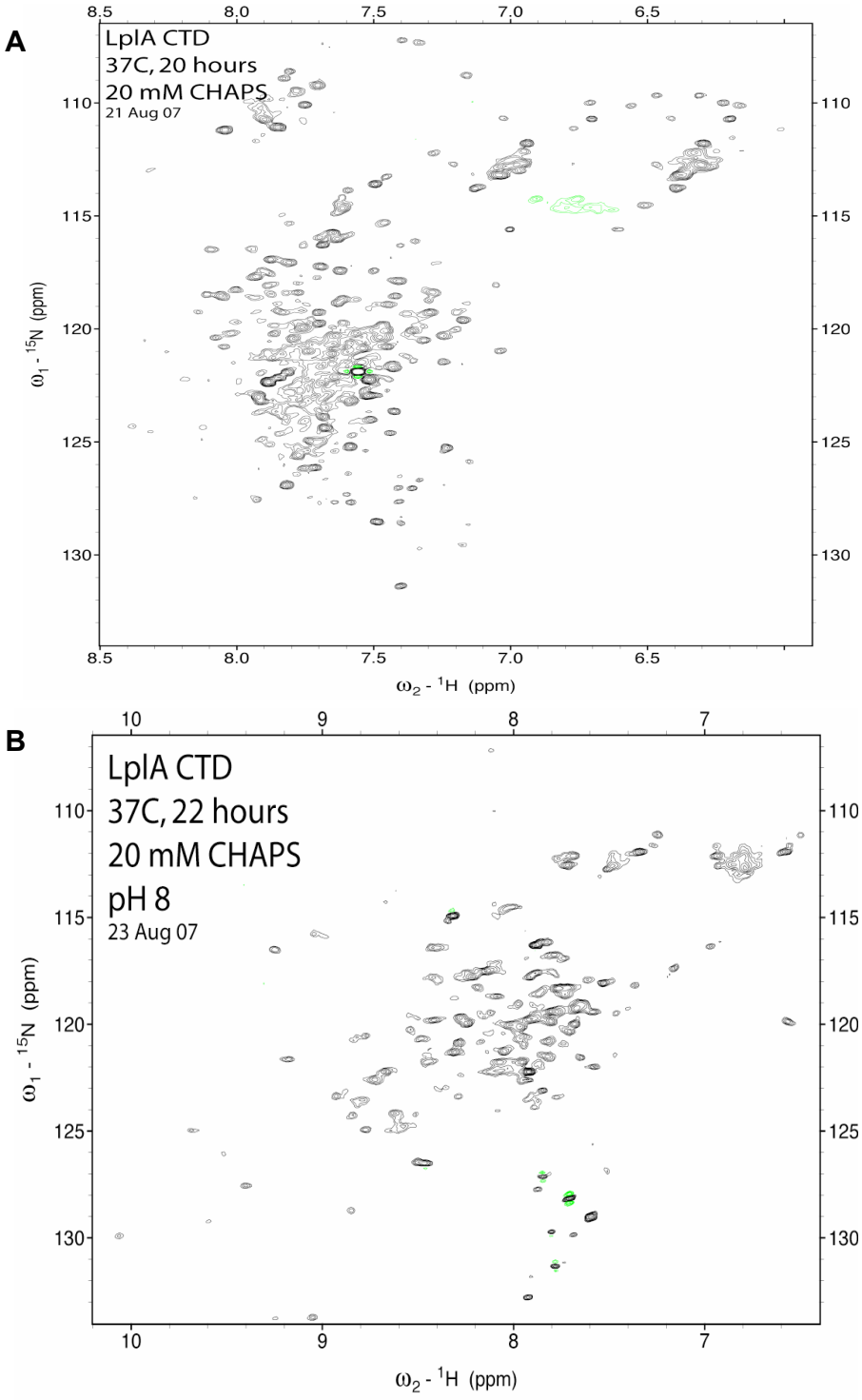


Fig. 4.10 HSQC spectra of His-tagged CTD in the presence of 20 mM CHAPS at 37°C. **A**: His-tagged CTD in 20 mM MES, pH 6.5, 100 mM NaCl. Number of scans: 272, proton spectral width: 8000 Hz, nitrogen spectral width: 1703 Hz. **B**: Analysis of His-CTD in 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM CHAPS. Number of scans: 304, proton spectral width: 8000 Hz, nitrogen spectral width: 1763 Hz.

Despite using different conditions, the signal dispersion and uniformity of His-CTD HSQC spectra did not improve sufficiently to permit detailed NMR studies. As the His-tag of the His-CTD constitutes approximately 18% of the total size of the protein, it was considered worthwhile to assess whether nontagged CTD will give better NMR signal dispersion as the His-tag may be affecting stability and folding of the protein, causing aggregation and/or precipitation. To generate sufficient amounts of ntCTD, the cell lysate was subjected to an ammonium sulphate cut with 1.5 M and 2.5 M concentrations (section 3.2.11). The 2.5 M supernatant was found to contain the ntCTD. As the theoretical pI of ntCTD is 5.58, ntCTD was buffer exchanged into 50 mM Tris-HCl, pH 7.5, and subjected to ion-exchange chromatography using a HiTrap Q FF column and eluted with a linear NaCl gradient of 0 - 0.5 M. This method resulted in 80% pure ntCTD for NMR analyses and the protein was dialyzed into 50 mM Tris-HCl, pH 7.5. HSQC spectra were recorded at 15°C, 25°C, 37°C and 50°C. The HSQC spectra recorded at all the temperatures showed very good signal dispersion with uniform peak intensity (Fig. 4.11), although the acquisition times were relatively long due to low sample concentration. The protein was expected to yield 85 backbone NH peaks (accounting for the three proline residues), but only ~ 55 peaks were observed. This can be attributed to low protein concentration or some parts of the protein being unstructured. Nonetheless, the lack of change with temperature in the HSQC showed that the ntCTD conformation was largely stable over a wide range of temperature, confirming its thermostability.

With little success in using NMR spectroscopy as a tool to look at the structure of CTD, attempts to crystallise the LplA-CTD complex were made by setting up crystal screens. Both sitting and hanging drop methods were employed and several known crystal screens have been set up. At the time of writing this thesis there was no sign of any favourable conditions for crystallisation of the complex and more conditions are being used.

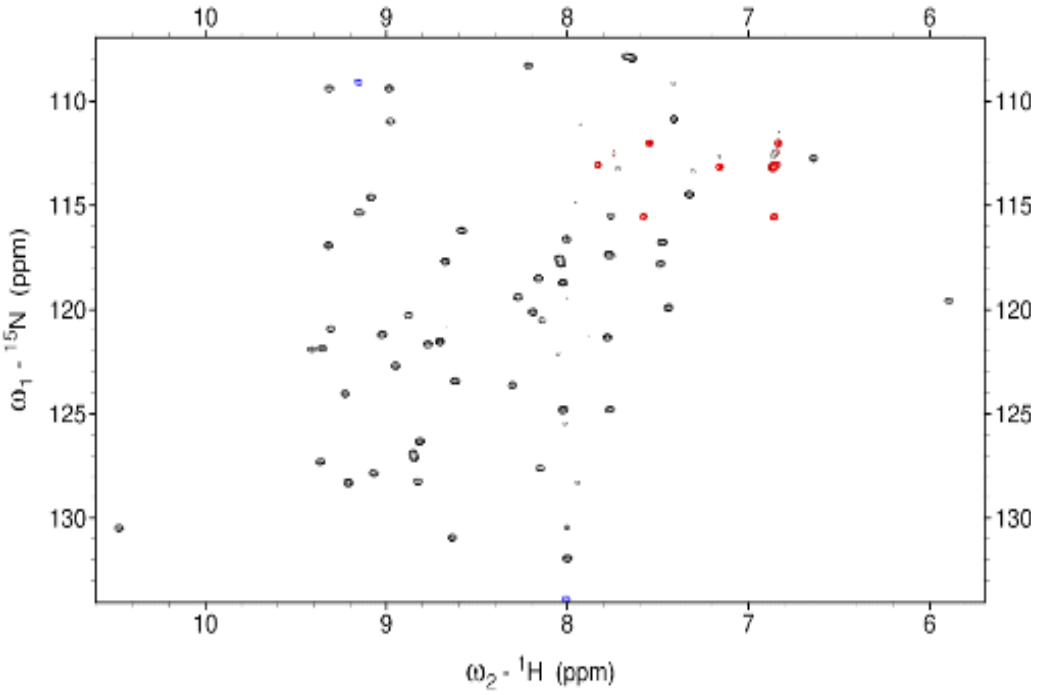


Fig. 4.11 HSQC spectrum of non-tagged CTD at 15°C. The spectrum was recorded for 21.5 h. Number of scans: 320, proton spectral width: 8000 Hz, nitrogen spectral width: 1702 Hz. Only about 55 backbone NH peaks are observed. Peaks in black and blue are from the backbone NH groups; the peak beyond 10 ppm corresponds to the single tryptophan residue. Red peaks are from asparagine and glutamine side chain NH_2 groups.

4.3 Discussion and conclusions

A major feature of the gene and protein sequence analysis is that genes *ta0514* and *ta0513m* are likely to represent the N- and C-terminal domain of a full-length *Tp. acidophilum* LplA, respectively. BLASTP searches indicate that this arrangement is present in other archaea as well, the most closely related case being in *Thermoplasma volcanium*, suggesting that this two-gene-arrangement may have a functional significance. Interestingly, this situation has been found in Euryarchaeota, whereas Crenarcheota have homologs of LipA and LipB or a full length LplA instead.

The two ORFs overlap by 1 base pair of the *ctd* termination codon (TGA) and the A of the initiation codon of *lpla* (ATG) and the translation initiation signals for the *lpla* fall within the *ctd* gene sequence. This is an interesting finding as the two gene products may reconstitute a full length lipoate protein ligase, although the genes are in reverse orientation to that seen in other lipoyl ligases. The 1

base pair overlap does not exclude independent transcription of the *lpla* gene. Programmed frameshifting has been reported in several other organisms (Giedroc and Cornish, 2008), including the archaeon *Sulfolobus solfataricus* (Cobucci-Ponzano *et al.*, 2006). The presence of a stem-loop structure covering the frameshift region as well as the unpaired A residue shared by the stop and start codon within this stem structure strongly indicates the possibility of a programmed frameshift (Giedroc and Cornish, 2008). At the transcriptional level, the formation of such a mRNA loop-structure could promote ribosome pausing and, together with the *lpla* SD sequence, allow programmed -1 frameshifting to give correct initiation of translation for both genes (Cobucci-Ponzano *et al.*, 2006). Such regulatory mechanism can be envisaged as transcription elongation is not a continuous process but pausing, stalling and arrest have been reported to occur frequently (Borukhov and Nudler, 2008). Moreover, if it is assumed that both gene products are required for full lipoylation activity, this scenario implies a level of regulation of how much active enzyme is present *in vivo*. If both proteins are required for full lipoate protein ligase activity, they must be individually stable *in vivo* and present in the correct stoichiometric ratio. A close, coordinated transcription process could just ensure this.

It can be concluded from the sequence analyses that both genes are potentially linked and controlled at the level of transcription and translation, which may help their association at the protein level. Nonetheless, without experimental data, it is very difficult to predict the transcriptional regulation and translation efficiency of both genes. More information can only be obtained from *in vitro* transcription and electromobility shift assays.

More work is needed to elucidate the structure of CTD, which may throw new light on protein- protein interactions between CTD and LplA and how they work in tandem to bring about the lipoylation activity. As the above analyses would predict that both, LplA and CTD are required for lipoylation, the next chapter describes the investigation of their lipoylation activity.

Chapter 5

Lipoate protein ligase activity in *Thermoplasma acidophilum*

5.1 Introduction

The gene and protein sequence analysis presented in the previous chapter suggests that LplA and CTD may be needed in tandem to reconstitute a functional lipoate protein ligase activity but no experimental evidence for this hypothesis has been presented as yet. The gel shift assay is a well-established qualitative assay for lipoylation and the E2lipD construct was carefully chosen as substrate to make sure that an acceptable assay could be established (Fujiwara *et al.*, 1995). In the current chapter, the results from a functional gel shift assay and Western blot analysis to detect lipoylation activity are described.

5.2 Results

5.2.1 Lipoylation activity of *Tp. acidophilum* LplA and CTD

The activity of the recombinant LplA and CTD was analyzed in a gel shift assay as described (section 3.2.15). As a positive control for lipoylation activity, *E. coli* LplA and *E. coli* E2lipD were incubated as described for the *Thermoplasma* proteins and lipoylation was assessed by a mobility shift of the *E. coli* E2lipD in non-denaturing PAGE. *Thermoplasma* E2lipD incubated with only LplA or CTD in the presence of LA and ATP did not show any change of its mobility on a non-denaturing PAGE gel. However, if LplA and CTD were incubated with E2lipD in the presence of LA and ATP, a clear downward shift of the E2lipD was detected (Fig 5.1 A, B). Both His-tagged and non-tagged constructs of LplA and CTD yielded similar results. Small amounts of a band corresponding to lipoylated E2lipD were present in the control reaction (Fig. 5.1 A, lane 5). LplA has been shown to convert LplA and ATP into lipoyl-AMP in the absence of CTD (Kim *et al.*, 2005). Therefore, some of the recombinant LplA may have already catalyzed the formation of lipoyl-AMP during expression in the *E. coli* host and/or the purification process. When mixed with CTD, lipoyl-AMP charged LplA is sufficient to carry out lipoylation of the E2lipD without the addition of ATP. Interestingly, *E. coli* LplA failed to lipoylate *Thermoplasma* E2lipD.

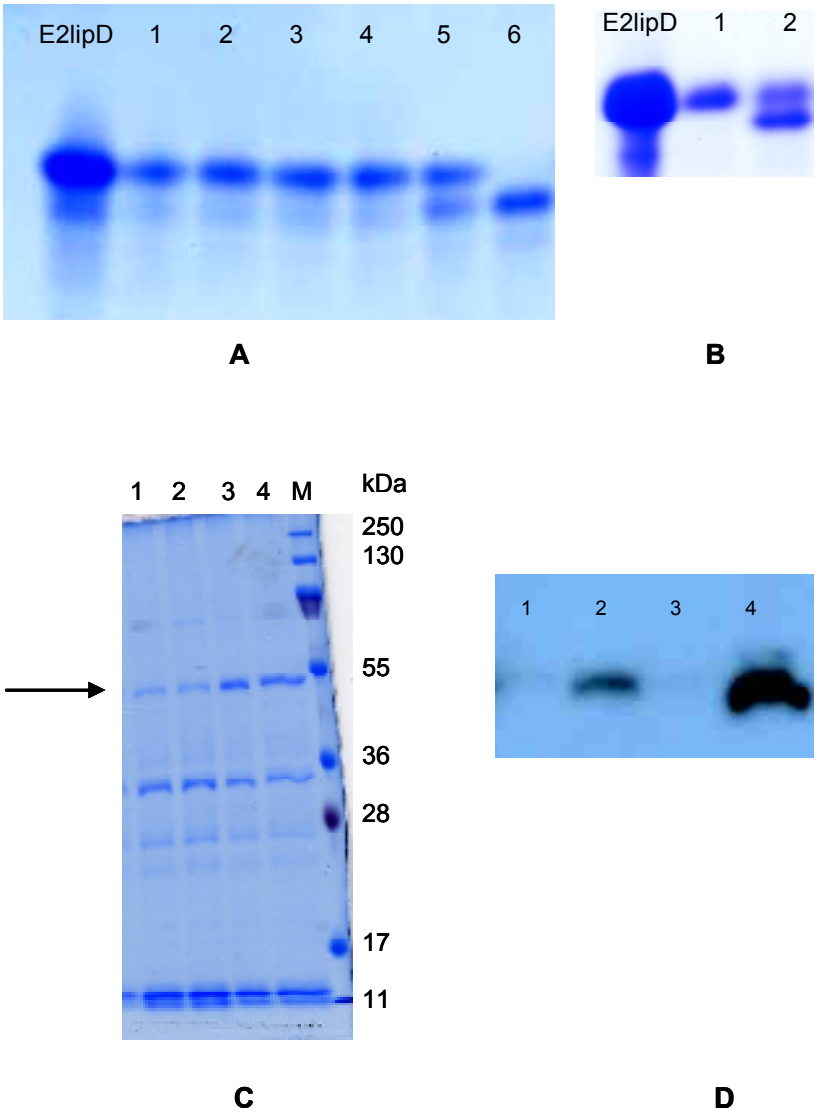


Fig. 5.1 Non-denaturing PAGE gels (A, B) showing lipoylation of E2lipD and SDS-PAGE gel (C) and Western blot (D) showing lipoylation of whole E2 by LplA and CTD. **A**: 250 µg of *Thermoplasma* E2lipD were used in the reaction (section 3.2.15). Non-lipoylated E2lipD was used as a negative control. Reactions contained 50 µg His-LplA (lane 1, 2) or 50 µg His-CTD (lane 3, 4) or both proteins (lane 5, 6). In lanes 1, 3 and 5 ATP was omitted; lane 2, 4 and 6 reactions contained ATP. **B**: Lipoylation activity of ntCTD (20 µg) and His-LplA (20 µg). The reaction contained 100 µg E2lipD. non-lipoylated E2lipD; lane 1, reaction contained no ATP; lane 2, reaction contained ATP. **C**: Lipoylation of whole E2 (50 µg). Lanes 1 and 3, reactions contained no ATP; lanes 2 and 4, reactions contained ATP. Reactions contained 20 µg CTD and 5 µg LplA (lane 1 and 2) or 40 µg CTD and 10 µg LplA (lane 3 and 4). Whole E2 indicated by arrow. **D**: Western blot of C using anti-lipoic acid antibody. Lanes 1 and 3, correspond to lanes 1 and 3 of C; lanes 2 and 4, correspond to lanes 2 and 4 of C.

It should be noted that the natural substrate of lipoate protein ligases is the E2 protein and attempts have been made to test the lipoylation of whole E2 in a gel shift assay. Whole E2 was not suitable for analysis by non-denaturing PAGE due to aggregation. Instead, the whole E2 was tested for lipoylation by running it on SDS-PAGE (Fig. 5.1, C) and the resulting gel was subjected to a Western blot using anti-lipoic acid antibody (section 3.2.17). This Western blot analysis showed that LplA and CTD lipoylated whole E2 as well (Fig. 5.1, D). It was also observed that the degree of lipoylation increased with the amount of enzyme used in the assay. For accuracy the experiment was reproduced three times.

5.2.2 Quantification of lipoylation using mass spectrometry

To quantitate the lipoylation of E2lipD, the lipoylated E2lipD was analysed for change in its total mass using quadrupole time of flight (Q-TOF) (section 3.2.18) (Chernushevich *et al.*, 2001). Based on a previous study (Heath *et al.*, 2007) in *Thermoplasma* only one site located in the E2lipD, K42, is expected to be lipoylated. MS analysis of unmodified E2lipD yielded a molecular mass of 11,297 Da (Fig. 5.2, A), which corresponded to the expected mass calculated from the amino acid sequence. Analysis of the lipoylated E2lipD yielded a molecular mass of 11485.15 Da, a difference of 188.06 Da from the unmodified E2lipD (Fig. 5.2, B). This change in molecular mass is in agreement with the expected result, if only one lipoyl-moiety (M = 188 Da) had been added to E2lipD.

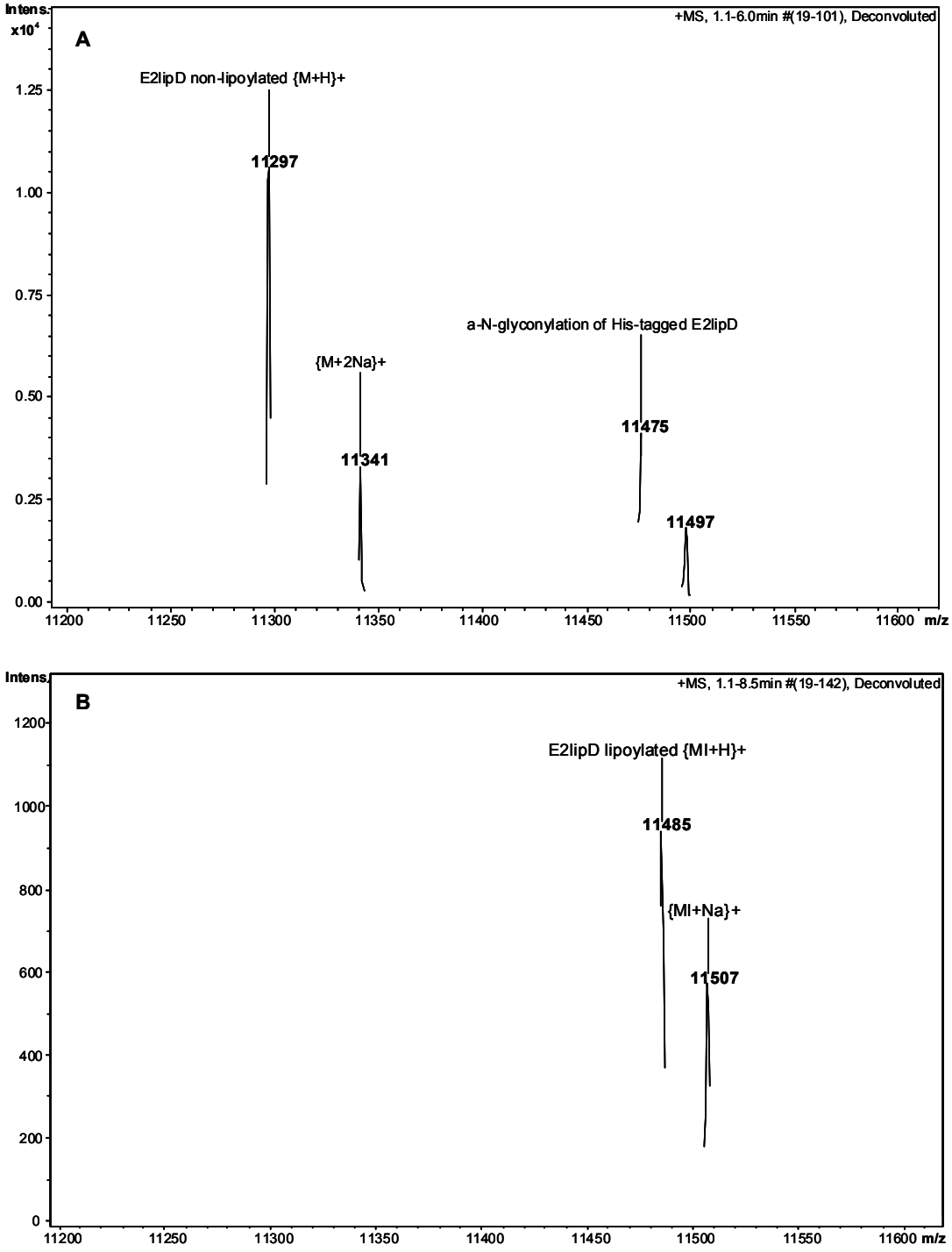


Fig. 5.2 Mass spectra (Q-TOF) of E2lipD, non-lipoylated (A) and lipoylated (B) (section 3.2.18).

5.2.3 Time course of lipoylation activity

After establishing that both LplA and CTD are required to carry out lipoylation, and that *Thermoplasma* E2lipD is lipoylated at only one site as expected, a study to analyse the enzyme activity as a function of time was carried out. The assay was set up as described previously (section 3.2.15) with His-LplA and His-CTD at 20°C (Fig. 5.3). A clear mobility shift for E2lipD was observed within 15 min. If ntLplA and ntCTD were used, the reaction required 10 h at 20°C before a shift was visible; a clear shift was only visible after 18 h of incubation. Furthermore, with non-tagged proteins, the reaction did not reach completion within the observation time of the experiment (Fig. 5.4).

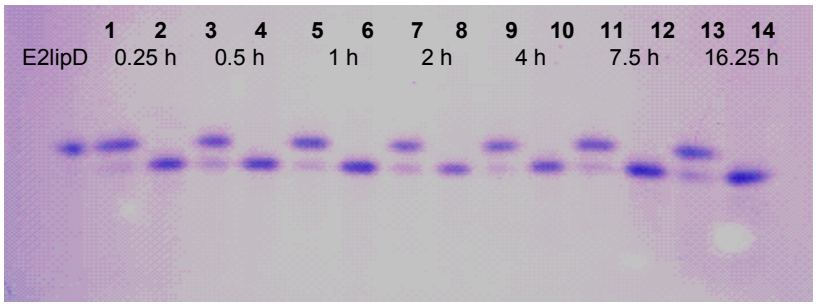


Fig. 5.3 Gel shift assay testing for lipoylation activity of His-LplA and His-CTD. Non-lipoylated E2lipD was used as a control. The reaction contained 100 µg E2lipD, 20 µg His-LplA and 20 µg His-CTD and LA; ATP was added to start the reaction (lanes 2, 4, 6, 8, 10, 12, 14). Lanes 1, 3, 5, 7, 9, 11 and 13 did not contain ATP. The reaction was incubated at 20°C and samples were taken at different time points.

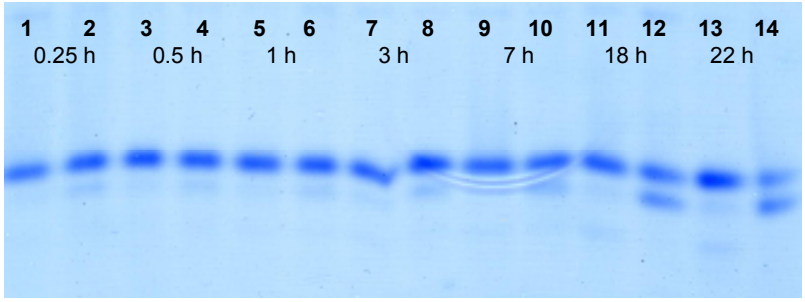


Fig. 5.4 Gel shift assay testing for lipoylation activity of ntLplA (20 µg) and ntCTD (20 µg). Each reaction contained 200 µg E2lipD. ATP was added to start the reaction (lanes 2, 4, 6, 8, 10, 12, 14). Lanes 1, 3, 5, 7, 9, 11 and 13 did not contain ATP. The reaction was incubated at 20°C and samples were taken at different time points.

5.2.4 Stoichiometry experiments

The optimum molar ratio of LplA to CTD was also tested using *in vitro* gel shift assays. The His-tagged constructs of the proteins were used for this purpose for ease of experiments. The molar ratios of LplA:CTD or CTD:LplA were 1:1, 1:2, 1:3 and 1:4 (Fig. 5.5). The experiment indicated that a 1:1 ratio was optimal for a complete shift of E2lipD under the given conditions, if 8 µM of the purified protein was used (Fig. 5.5, lane 10).

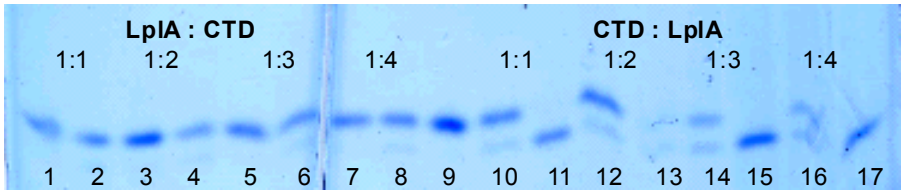


Fig. 5.5 Stoichiometry experiments with His-LplA and His-CTD. The lipoylation assay contained 50 µg E2lipD. Lane 9, non-lipoylated E2lipD. Reactions in lanes 1, 3, 5, 7, 10, 12, 14 and 16 contained no ATP. Reactions in lanes 2, 4, 6, 8, 11, 13, 15 and 17 contained ATP.

5.2.5 Lipoylation activity with the intermediate lipoyl-AMP

Lipoyl-AMP (Fig. 5.6) is the reaction intermediate between the reactant lipoic acid and the lipoyl moiety (Fig. 1.5) that is covalently attached to the lysine residue on the E2 lipoyl domain (Cronan *et al.*, 2005). The lipoylation reaction is usually carried out by a single polypeptide, lipoate protein ligase (Cronan *et al.*, 2005). In the case of *Tp. acidophilum*, as discussed above, the lipoate protein ligase activity is constituted by two individual proteins. A crystal structure of *Tp.* LplA with the intermediate lipoyl-AMP has been described and has been interpreted as a sign of partial enzyme activity (Kim *et al.*, 2005). In the light of our results, it is possible to hypothesise the functional roles of the two proteins, where *Tp.* LplA may be catalysing the formation of lipoyl-AMP and CTD may be responsible for carrying out the transfer of this intermediate to the acceptor protein (McManus *et al.*, 2006). To test this hypothesis, the substrate intermediate lipoyl-AMP was synthesised (section 3.2.16) and used instead of LA in the gel shift assay. The final synthesis product was characterized using thin-layer chromatography (TLC), infrared spectroscopy (IR) and nuclear magnetic resonance spectroscopy (NMR).

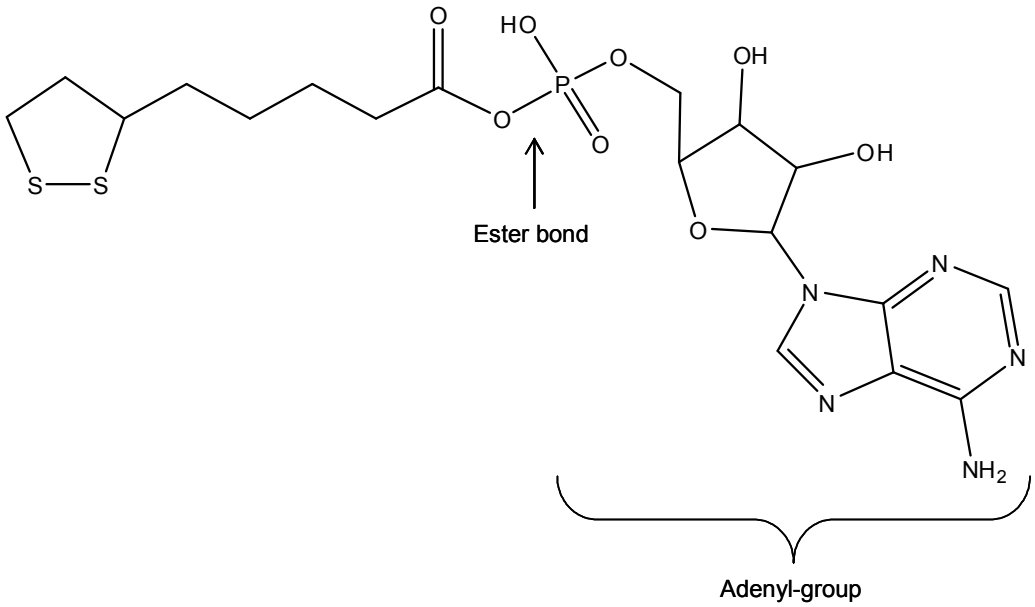


Fig. 5.6 Structure of lipoyl-AMP. The ester bond between lipoic acid and AMP is indicated by the arrow.

5.2.5.1 Thin-layer chromatography of the lipoyl-AMP synthesis product

TLC is a type of adsorption chromatography routinely used to analyse organic molecules. The separation takes place due to the difference in the relative mobility of different compounds in a given solvent (mobile phase) when silica is used as the stationary phase (Touchstone, 1978). TLC is regularly employed as a first technique to check the progress and completion of a chemical reaction as the reactants and products are expected to have different relative mobilities in the same mobile phase.

Fig. 5.7 shows the results of the TLC analysis of the lipoyl-AMP synthesis. The lipoyl anhydride (Fig. 5.7, lane 1) shows the same relative mobility as LA (Fig. 5.7, lane 3) on the TLC. This is due to the conversion of lipoyl anhydride back to LA by reacting with moisture on the silica. Fig. 5.7, lane 5 shows a species with a different relative mobility than any of the starting materials and synthesis intermediates thereby indicating the presence of a synthesis product that is expected to be lipoyl-AMP. As it cannot be concluded from TLC alone that this synthesis product is lipoyl-AMP, its identity was further analysed.

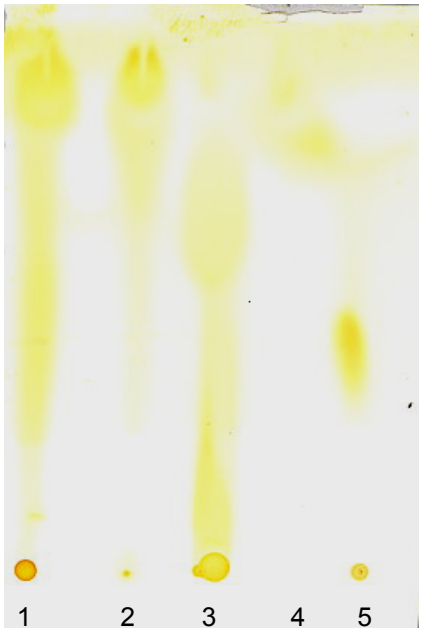


Fig. 5.7 Thin-layer chromatography of synthesis components, synthesis intermediate and lipoyl-AMP. Samples from the different synthesis steps were run on the TLC in methanol-chloroform (1:1 v/v). Lane 1, lipoyl anhydride; lane 2, DCC; lane 3, LA; lane 4, AMP; lane 5, lipoyl-AMP.

5.2.5.2 Infrared spectroscopy of the lipoyl-AMP synthesis product

The TLC analysis was complemented by infrared (IR) spectroscopy as this technique can give information about the functional groups that are present in a molecule. Organic molecules are not solid structures; bonds between atoms are vibrating as they constantly change their length. If molecules are exposed to electromagnetic radiation, different functional groups will absorb energy at a frequency that matches their frequency vibration; hence a change of functional groups can be detected using IR. In lipoyl-AMP, LA and AMP are linked through an ester bond, which has a different predicted characteristic absorption frequency than either the AMP or the LA. The IR spectra of LA and lipoyl-AMP are shown in Fig. 5.8; an IR spectrum of AMP is shown in Fig. 5.9. An IR spectrum can be divided into two main regions: the finger print region, which is distinct for every compound (1500 – 600 cm^{-1}) and the group frequency region (4000 – 1500 cm^{-1}) in which functional groups have specific absorption bands (McMurry,1999; Harwood *et al.*, 1999).

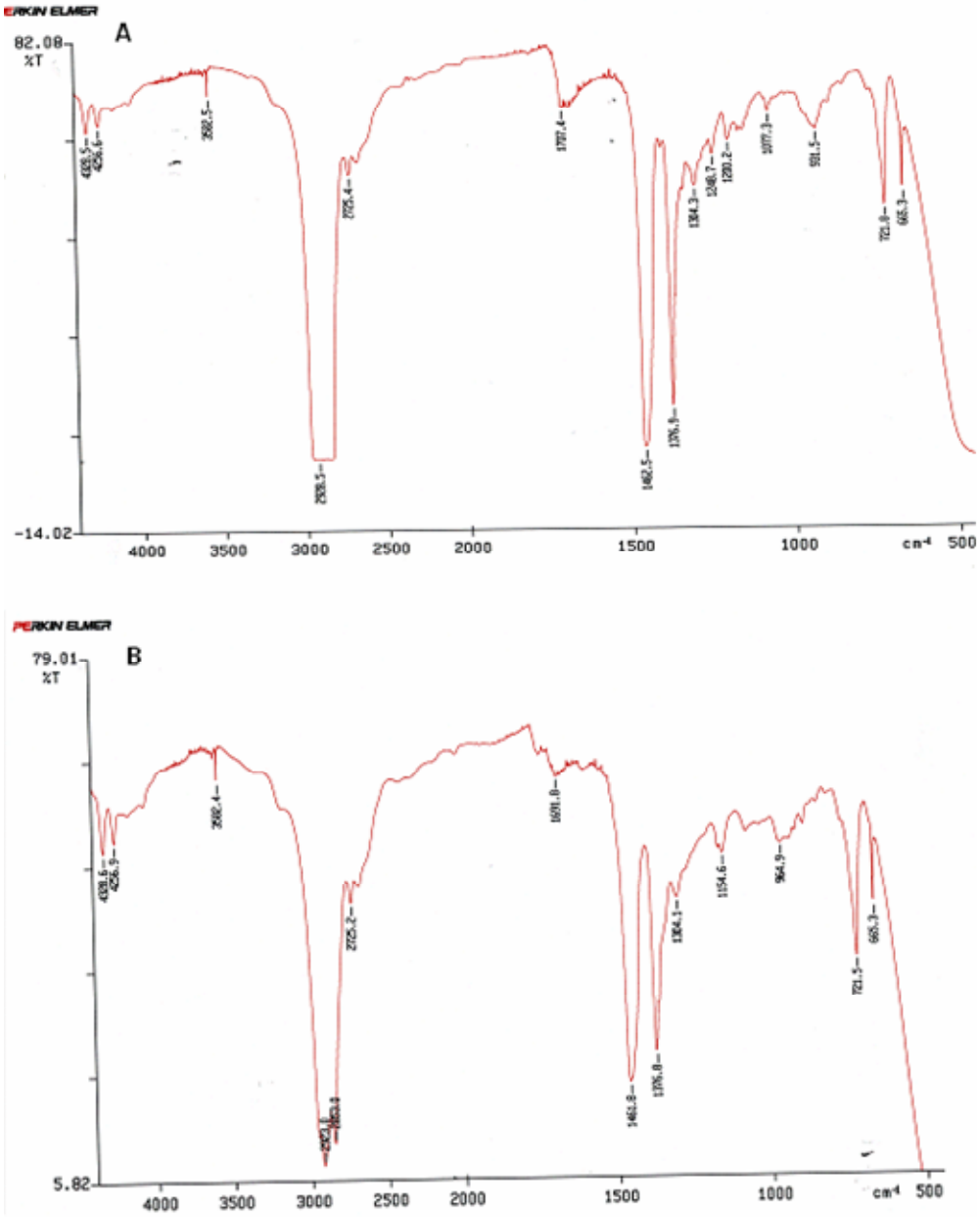


Fig. 5.8 IR spectrum of LA (A) and lipoyl-AMP (B). The solid synthesis product was prepared by mulling with Nujol oil and the resulting viscous mixture was analysed using KBr plates.

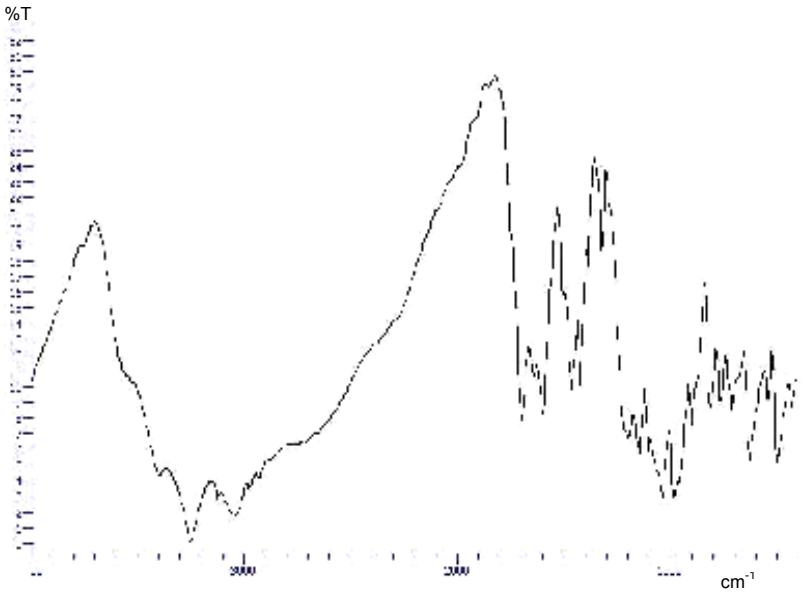


Fig. 5.9 IR spectrum of adenosine monophosphate (AMP) (taken from Chemexper.com database).

A comparison of the different IR spectra showed clear differences, especially in the fingerprint region. As fingerprint regions are specific to compounds, this indicated that the synthesis product was not LA. The lipoyl-AMP spectrum showed absorptions at 1154.6 cm^{-1} and 964.9 cm^{-1} that are absent from the LA spectrum. The characteristic IR absorption of carbonyl compounds (carbon atom double-bonded to oxygen atom) is located at 1670 – 1780 cm^{-1} in IR spectra (McMurry, 1999). In the case of LA, it is a carboxylic acid group, whereas lipoyl-AMP is a carbonyl ester. The LA IR spectrum showed a clear signal at 1707.4 cm^{-1} whereas the lipoyl-AMP spectrum showed a signal at 1691.8 cm^{-1}. This shift in wavelength can be associated with lower bond energy of the phosphoester bond in lipoyl-AMP. Moreover, the obtained spectra are different from the AMP spectrum (Fig. 5.9).

5.2.5.3 NMR analysis of the lipoyl-AMP synthesis product

As a complementary test, NMR spectroscopy was carried out on lipoyl-AMP. NMR spectra for AMP (Appendix I) and LA (Appendix II) confirmed the purity of the starting material and provided reference spectra for comparison with the spectra of lipoyl-AMP. A ^{31}P NMR spectrum of AMP gave only one signal at 0 ppm, whereas the signal in the ^{31}P spectrum of lipoyl-AMP was seen shifted

upfield to -1 ppm. This was expected due to the inductive effect of the attached lipoic acid part (Fig. 5.10). The ^{1}H NMR spectrum of lipoyl-AMP corresponded to a combination of the AMP and LA spectra (Fig. 5.11). The critical proton signals from the different parts of the lipoyl-AMP molecule matched with the predicted values (ChemNMR, CambridgeSoft®), thereby providing further evidence for a successful synthesis (Fig. 5.11).

5.2.5.4 Gel shift assay using the synthetic lipoyl-AMP

Lipoyl-AMP was included in a gel shift assay replacing LA and ATP. Despite several attempts, no clear results could be obtained as the protein samples seemed to precipitate in the wells. This result could be due to the presence of solvent contaminants in lipoyl-AMP. However, further attempts to avoid the use of such chemicals in the synthetic process by modifying the synthesis protocol using benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBop) did not yield the desired product for the gel shift assay.

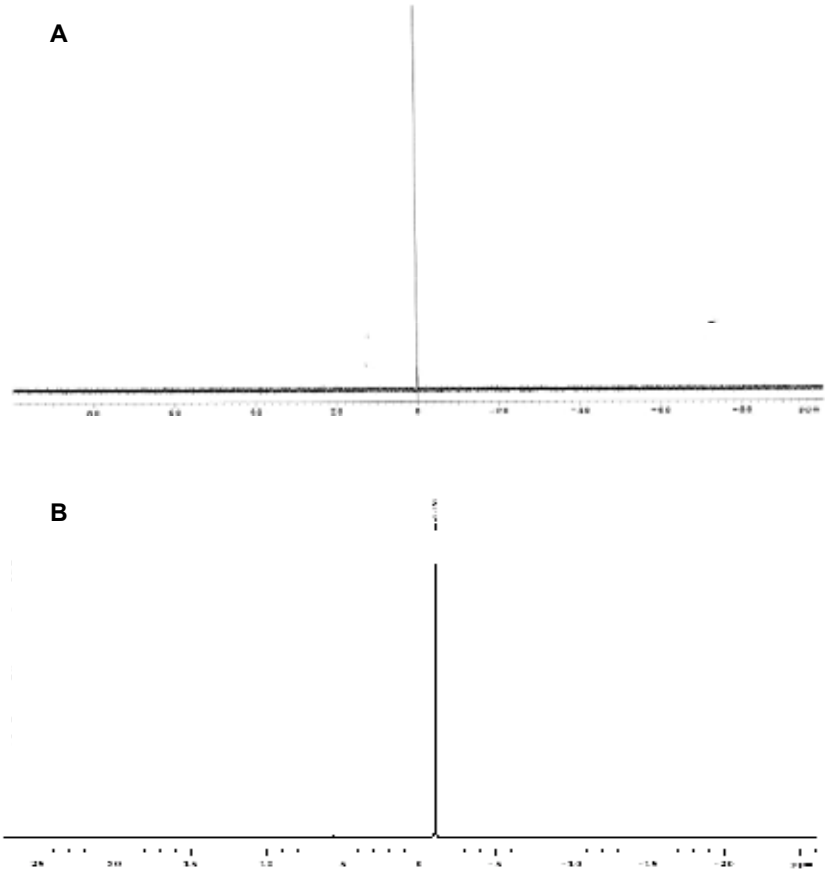


Fig. 5.10 ^{31}P NMR spectra of AMP (A) and lipoyl-AMP (B).

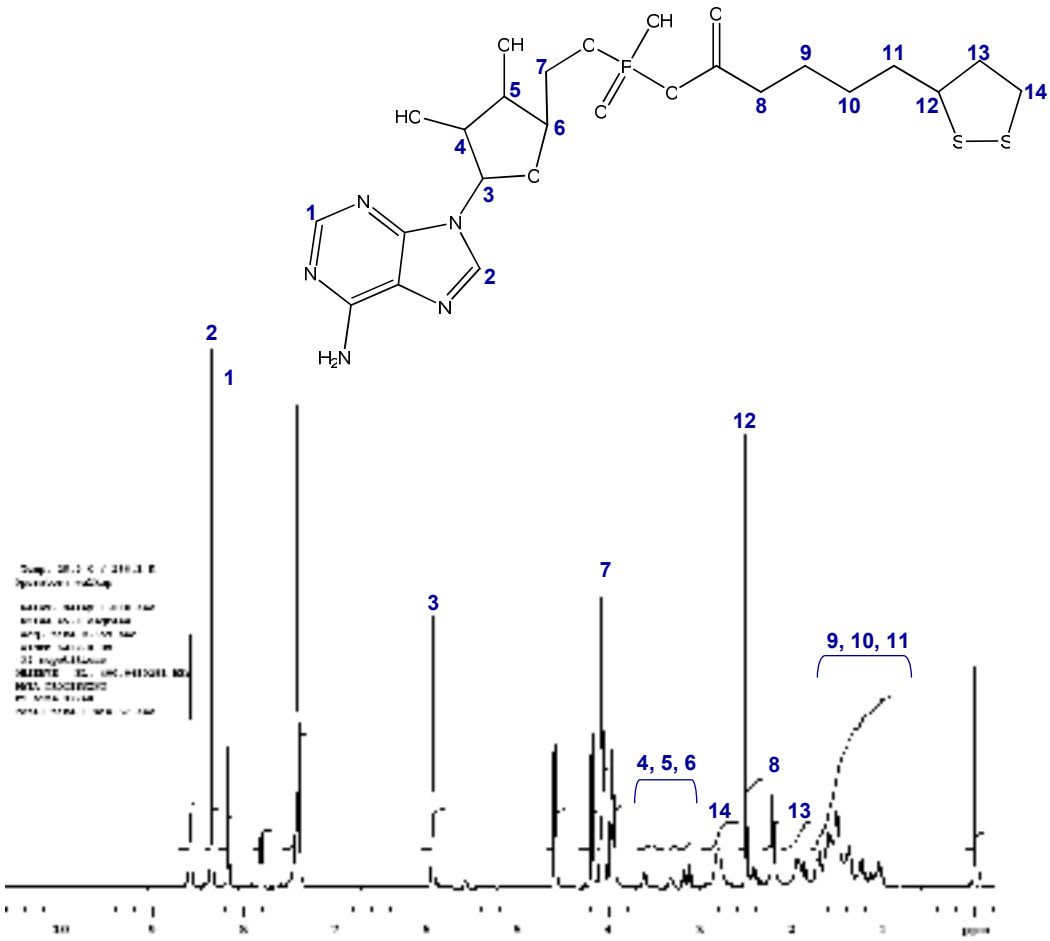


Fig. 5.11 ^{1}H NMR spectrum of lipoyl-AMP. The protons in the lipoyl-AMP molecule and their corresponding signal in the NMR spectrum are indicated by the numbering.

5.3 Discussion and conclusions

The lipoylation gel shift assay presented in this chapter is the first report of an active lipoate protein ligase from an archaeon. Also, it is a first report of a lipoate protein ligase that requires two individual proteins to constitute a functional enzyme. The stoichiometry for constituting an active enzyme worked out to a 1:1 molar ratio of CTD to LplA. The activity assays were carried out at 20°C, which is 35°C lower than the optimum growth temperature of *Tp. acidophilum*. The effect of temperature on the reaction rate (k) can be described by the Arrhenius equation: $k = A\, e^{(-Ea / RT)}$, where A is a constant, Ea is the Arrhenius activation energy, R is the universal gas constant and T is the temperature in Kelvin. Based on this equation, even a small change in temperature has a significant effect on the overall rate of reaction. As a rule of thumb, every

10°C increase in reaction temperature doubles the rate constant in the case of many enzymes (Vieille and Zeikus, 2001). Based on this the time it took for lipoylation at 20°C is not unexpected if LplA and CTD were designed to be active at 55°C.

The role of the individual proteins in the reaction mechanism using lipoyl-AMP could not be established due to some unknown reasons possibly relating to the solvent impurities in synthetic lipoyl-AMP. If LplA and CTD associate forming an overall lipoate protein ligase structure in a 1:1 stoichiometry, the active site may be formed by LplA and CTD. Also CTD may act as a modifier of LplA catalytic activity, changing LplA interactions with the substrate/intermediate thereby allowing the transfer to take place. Such a scenario is not unlikely and explains the requirement of both the proteins for carrying out lipoylation, as well as the presence of lipoyl-AMP in the LplA crystal structure (Kim *et al.*, 2005). Glutamate-cysteine ligase is an example of a heterodimer made up of a catalytic and modifier domain (Lee *et al.*, 2006) and *Tp. acidophilum* lipoyl protein ligase could be another one. In order to establish a possible interaction between the LplA, CTD and E2lipD, some biophysical studies were needed and are described in the next chapter.

CHAPTER 6

Biophysical analysis of protein interactions between *Tp. acidophilum* LplA, CTD and E2lipD

6.1 Introduction

After demonstrating that both LplA and CTD are required to carry out the lipoylation reaction, the next step was to study the nature of interaction between these proteins. In order to study the physical nature of the interaction a simple pull-down assay and gel permeation chromatography (GPC) were employed. Dynamic light scattering (DLS) was used to determine the size of any resultant assembly out of the interaction of LplA and CTD. Surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) were used to study the thermodynamics of the process.

6.2 Results

6.2.1 Pull down assays

As per the procedure described in section 3.3.1, His-LplA was immobilized on a Ni^{2+} column as the bait and the column was washed to remove excess His-LplA with 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. ntCTD was passed over the same column in the same buffer and the column was extensively washed to remove any unbound ntCTD. His-LplA and ntCTD were eluted together from the column by applying a linear gradient of imidazole (0 – 0.5 M) (Fig. 6.1, lane 9). Bands different to the ones corresponding to His-LplA and ntCTD can be explained by the presence of proteins that were already present in the initial ntCTD sample. This His-LplA-ntCTD complex was found to be active when tested for lipoylation activity in a gel shift assay (Fig. 6.2).

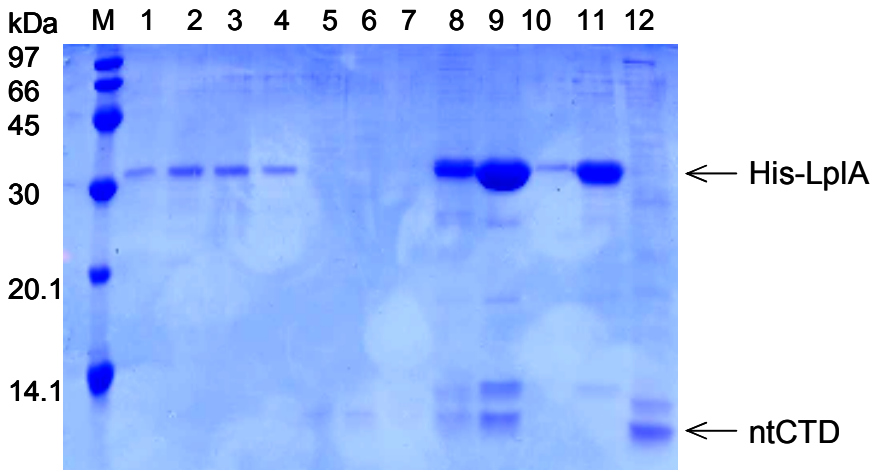


Fig. 6.1 SDS-PAGE gel of the pull-down assay (section 3.3.1) using His-LplA as bait and ntCTD as prey. M, protein marker (kDa); lanes 1 – 4, flowthrough of the His-LplA sample which has been injected over a 1 ml Ni^{2+} column and washes; lanes 5 – 7, flow-through of the ntCTD sample which has been injected over a 1 ml Ni^{2+} column charged with His-LplA; lanes 8 – 10, fractions eluted with imidazole; lane 11, His-LplA; lane 12, ntCTD. His-LplA and ntCTD are indicated by arrows.

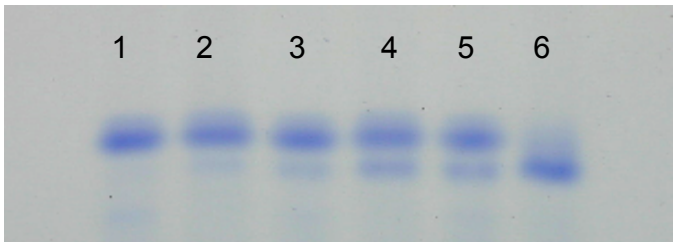


Fig 6.2. Non-denaturing PAGE gel showing lipoylation acitivty of the His-LplA-ntCTD complex from the pull-down assay (Fig. 6.1). The reaction mixture contained 250 µg E2lipD and was allowed to proceed at room temperature overnight (section 3.2.15). Different volumes of the complex mixture were added: 2 µl (lane 1, 2); 4 µl (lane 3, 4), 10 µl (lane 5, 6). ATP was omitted in lane 1, 3 and 5, whereas all assay components were present in lane 2, 4 and 6.

As the substrate of the lipoylation reaction is also a protein, the E2lipD, the possible interaction between LplA and CTD with E2lipD was investigated in another set of pull-down assays. If His-E2lipD was used as the bait, it was possible to retain both ntLplA and ntCTD separately (Fig. 6.3). Using approximately 75% pure recombinant ntLplA or ntCTD in the pull-down assay showed that the interaction between E2lipD and ntLplA (Fig. 6.3 A, lane 12) and between E2lipD and ntCTD (Fig. 6.3 B, lane 8) is specific.

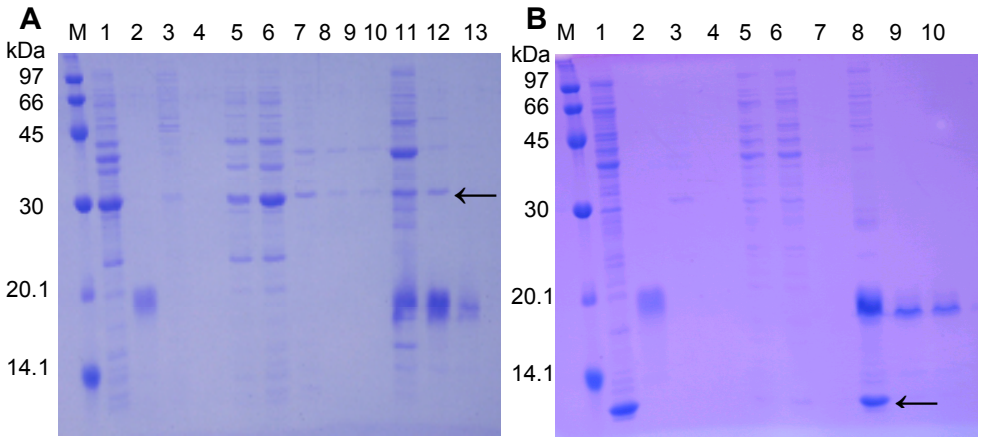


Fig. 6.3 SDS-PAGE gels of the pull-down assays (section 3.3.1) using His-E2lipD as bait and ntLplA (A) and ntCTD (B) as prey. M, protein marker (kDa). **A**: lane 1, ntLplA; lane 2, E2lipD; lanes 3 – 6, flowthrough of the ntLplA sample which has been injected over a 1 ml Ni^{2+} column charged with His-E2lipD; lanes 7 – 10, column wash; lanes 11 – 13, fractions eluted with imidazole. Arrow indicates ntLplA. **B**: lane 1, ntCTD; lane 2, E2lipD; lanes 3 and 4, flowthrough of the ntCTD sample which has been injected over the His-E2lipD charged Ni^{2+} column; lanes 5 – 7, wash of the column with buffer; lanes 8 – 10, fractions eluted with imidazole. Arrow indicates ntCTD.

As ntLplA and ntCTD had previously been successfully retained in a pull-down assay by His-E2lipD (Fig. 6.3), all three proteins were included in a modified pull-down experiment to try and understand the nature of any multi-protein complex they may be forming. The non-tagged versions of LplA and CTD have previously shown a tendency to retain co-purified proteins. To reduce the number of co-purified proteins during the experiment, it was decided to express, purify and prepare ntLplA, His-CTD and His-E2lipD (section 3.2) and His-CTD was treated with thrombin to cleave off the His-tag. Attempts to cleave off the His-tag from His-LplA using enteropeptidase proved difficult, as the yields were poor, grossly insufficient for pull-down assays. All proteins were buffer exchanged into 20 mM Tris-HCl, pH 7.5. His-E2lipD (30 mg/ml) was used as bait on a 1 ml HisTrap HP column. A mixture of ntLplA and ntCTD (1:1 equal concentration) was then used as prey and loaded onto the column in the running buffer (20 mM Tris-HCl, 0.3 M NaCl, 20 mM imidazole, pH 8.0). The column was extensively washed with running buffer to remove any non-specifically bound proteins. The ternary complex, constituted of His-E2lipD, ntLplA and ntCTD, was eluted with a linear gradient of imidazole (0 – 0.5 M) in the running buffer (Fig. 6.4, lane 7).

Initial Coomassie stained SDS-PAGE gels of the pull-down were difficult to analyse due to the tendency of E2lipD to smear. Therefore, less amount of protein was analysed on SDS-PAGE gels using silverstain (Fig. 6.4). Silverstaining showed that E2lipD was present in excess and that the eluted fraction containing the majority of E2lipD also contained ntLplA and ntCTD (Fig. 6.4, lane 7). Therefore, an interaction between the three proteins appears to exist even in the absence of the substrate, LA. The silverstain method also showed impurities in the ntLplA sample not visible on Coomassie stained gels. However, the fraction containing ntLplA retained by E2lipD did not contain any of these impurities. This indicated a specific interaction of LplA and E2lipD and that the co-purified proteins from the ntLplA sample may not be cross-interacting with E2lipD, LplA and/or CTD.

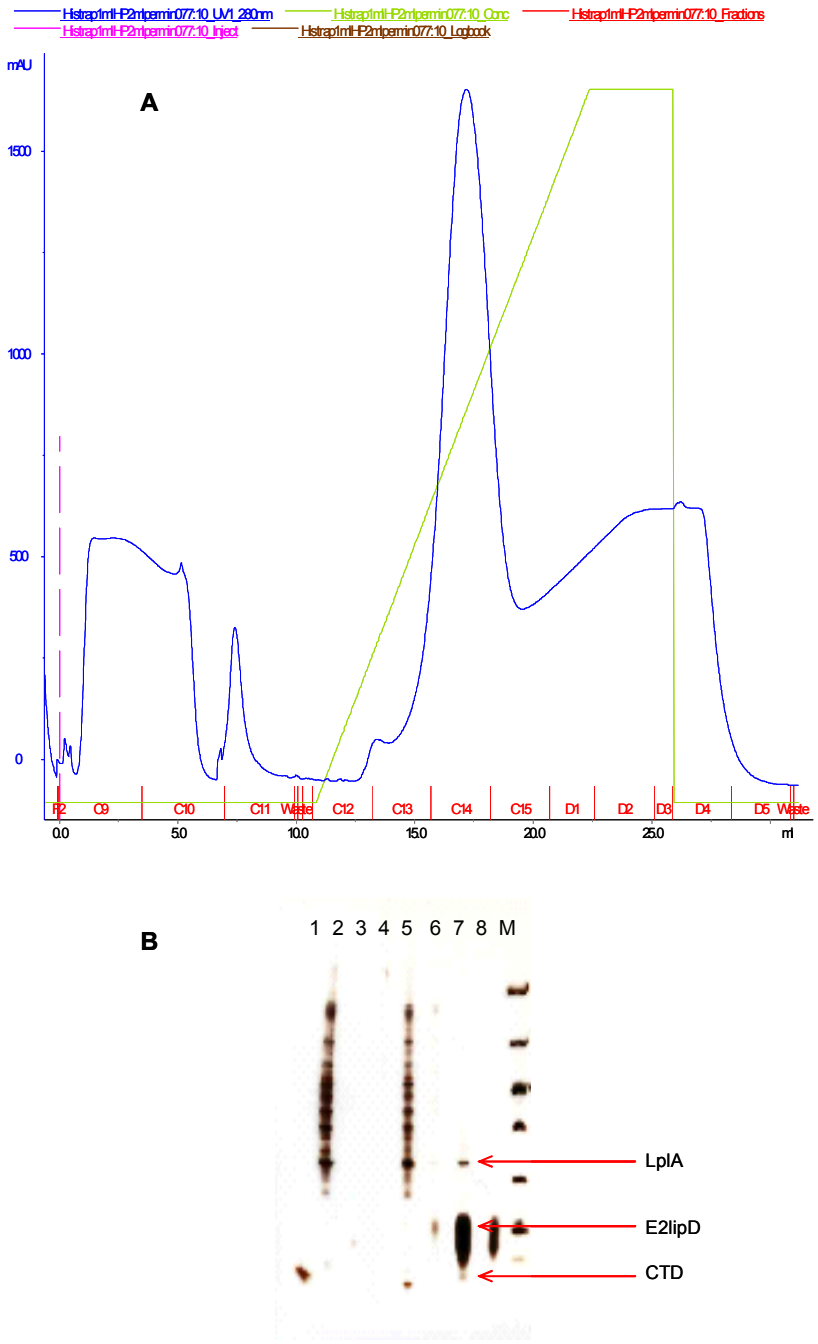


Fig. 6.4. Pull-down assay of a ternary E2lipD-LplA-CTD complex. **A**: Chromatogram of the pull-down showing the change in absorbance at A_{280nm} versus elution volume. A HisTrap HP column (1 ml) was saturated with His-E2lipD (30 mg/ml). Then a 1:1 mixture of ntLplA and ntCTD (equal concentration) was loaded and the column was washed with running buffer (20 mM Tris-HCl, 0.3 M NaCl, 20 mM imidazole, pH 8.0) until the A_{280nm} reading reached baseline. A linear gradient of imidazole (0 – 0.5 M) was applied in the running buffer to elute His-E2lipD from the column. Fractions C9 to C11 correspond to excess unbound LplA/CTD eluting from the column. With increasing imidazole concentration (green), protein is eluted off the column; majority of protein eluted in fraction C14. **B**: SDS-PAGE gel (silverstain) of fractions from the pull-down. Lane 1, ntCTD; lane 2, ntLplA; lanes 3 and 4, unbound excess His-E2lipD; lane 5, fraction C9; lane 6, fraction C13; 7, fraction C14; 8, fraction C15; M, protein marker (kDa). LplA, E2lipD and CTD are indicated by arrows.

6.2.2 Gel permeation chromatography – evidence for a ternary E2lipD-LplA-CTD complex

The pull-down assays indicated the possibility of a complex forming between His-LplA and ntCTD, His-E2lipD and ntLplA, His-E2lipD and ntCTD and between all three proteins. As there is the possibility that there was a heterogeneous mixture of His-E2lipD bound to either ntLplA or ntCTD, which would elute together, the presence of this possible ternary complex was tested using gel permeation chromatography (GPC). To study the ternary complex, fixed amounts of His-E2lipD (3 mg/ml), His-LplA (3 mg/ml) and His-CTD (1 mg/ml) in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, were mixed in a 1:1:1 molar ratio yielding a total of 3 mg protein. This mixture was applied to a pre-equilibrated Superdex 200 GPC column and eluted in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl as described in section 3.3.5. Under these conditions an elution profile with two distinct peaks was obtained (Fig. 6.5, A). SDS-PAGE revealed that the peaks correspond to complexes of His-LplA with His-CTD (Fig. 6.5, B lane 1) and His-LplA with His-E2lipD (Fig. 6.5, B lane 2). The molecular size estimates from GPC corresponded to a 1:1 ratio of His-LplA to His-CTD and His-LplA to His-E2lipD (Table 6.1). The experiment was repeated as described above, omitting NaCl from the running buffer. In the absence of NaCl, the GPC profile had only one major peak (Fig. 6.6, A). SDS-PAGE analysis of this peak revealed the presence of all the three proteins, His-LplA, His-CTD and His-E2lipD (Fig 6.6, B lane 3). Therefore, the presence or absence of NaCl seemed to have an effect on the formation of the protein complex. The observed effect of NaCl was not a single event and GPC data were reproducible over three runs. The molecular size estimates of the protein complexes, as per GPC, are summarized in Table 6.1.

Fractions D4 to D5 were concentrated using Amicon Ultra-15 centrifugal filter devices (Millipore) prior to SDS-Page gel analysis. The liquid flowing through the concentrator (filtrate) did not contain any protein, but all proteins, including the relatively small His-CTD, were retained in the concentrate. Washes of the concentrator showed that the proteins did not get stuck to the filter membrane (Fig. 6.6, C).

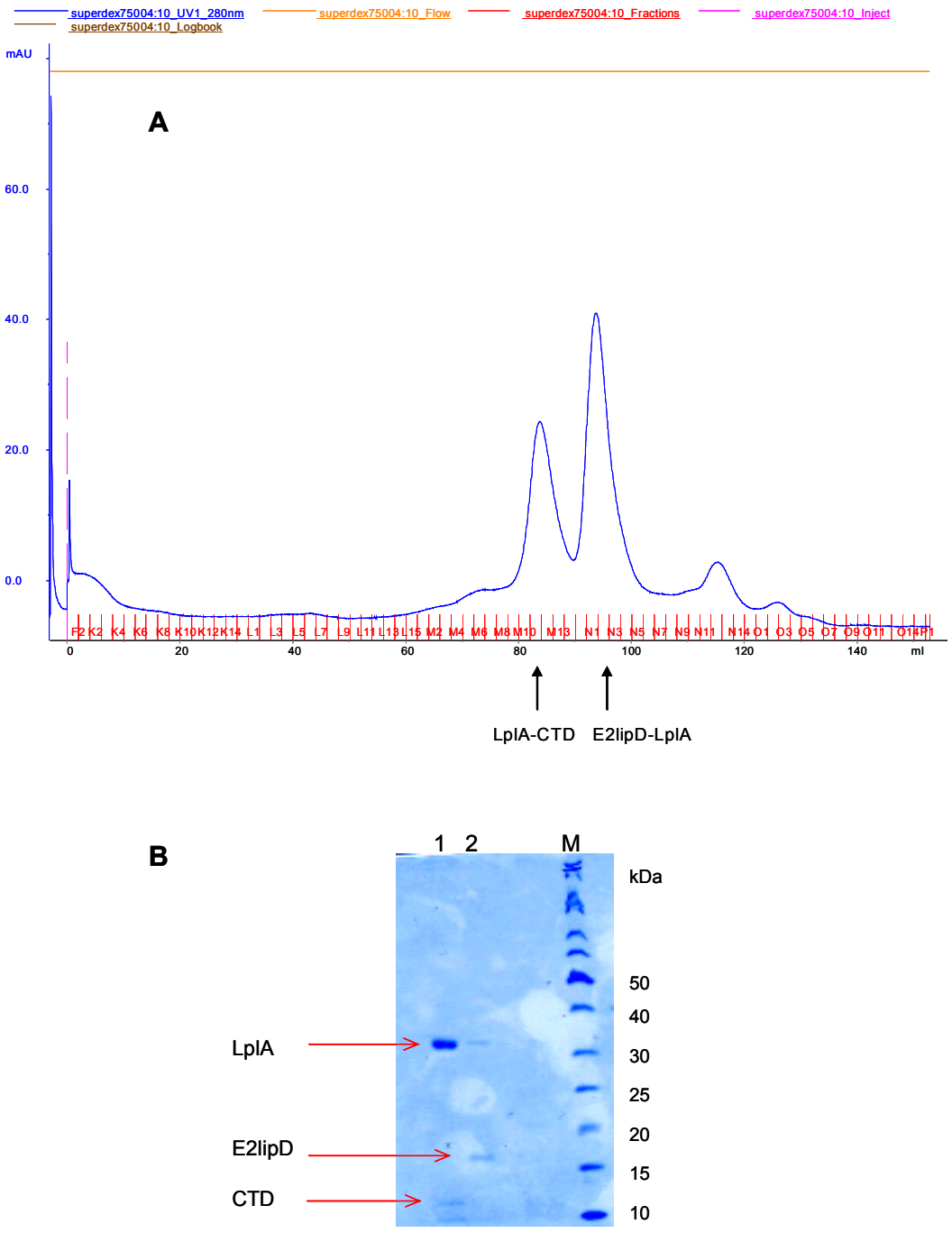


Fig. 6.5 GPC of His-E2lipD, His-LplA and His-CTD and SDS-PAGE analysis. **A**: GPC chromatogram. His-E2lipD, His-LplA and His-CTD in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, were mixed in a 1:1:1 molar ratio. This mixture was applied to a pre-equilibrated Superdex 200 GPC column and eluted in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl as described in section 3.3.5. The A_{280nm} trace recorded two main peaks in the elution profile. **B**: SDS-PAGE analysis of the protein peaks from GPC of a mixture of His-E2lipD, His-LplA and His-CTD. M, protein marker (kDa); lane 1, first peak: His-LplA and His-CTD; lane 2, second peak: His-LplA and His-E2lipD. His-LplA, His-CTD and His-E2lipD are indicated by arrows.

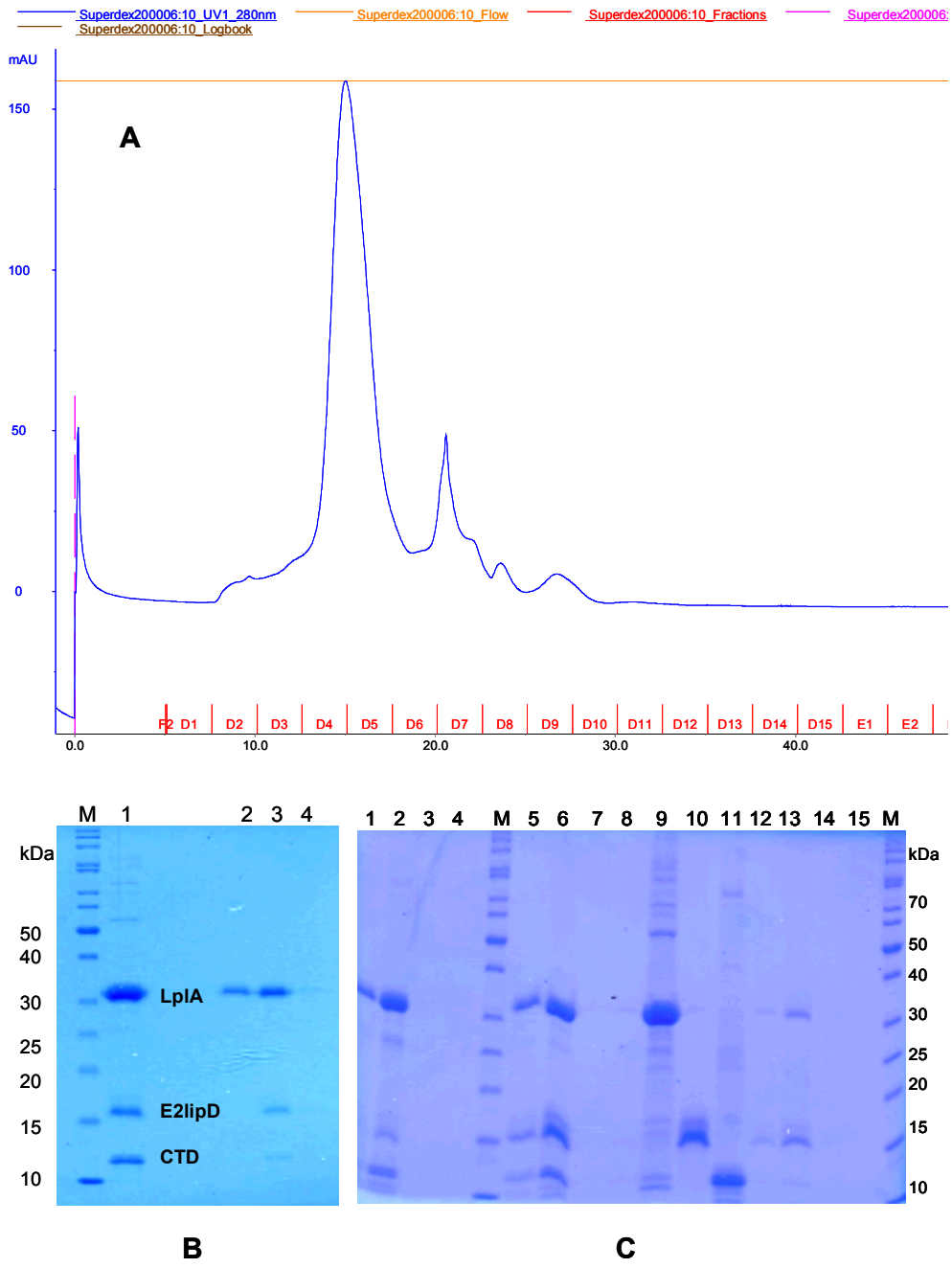


Fig. 6.6 GPC of His-E2lipD, His-LplA and His-CTD in the absence of NaCl (A) and SDS-PAGE gels of eluted fractions (B, C). M, protein marker (kDa) **A**: His-E2lipD, His-LplA and His-CTD in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, were mixed in a 1:1:1 molar ratio. This mixture was applied to a pre-equilibrated Superdex 200 GPC column and eluted in 20 mM Tris-HCl, pH 7.5, omitting NaCl. The A_{280nm} trace recorded one main peak in the elution profile. **B:** Lane 1, protein mixture before loading onto the GPC column; lanes 2 to 4 correspond to fractions D4 to D6, respectively. **C**: Concentrating GPC fractions (section 3.2.4). Lane 1, D4 before concentrating; lane 2, concentrated D4; lane 5, D5 before concentrating; lane 6, concentrated D5; lane 9, His-tagged LplA; lane 10, His-tagged E2lipD; lane 11, His-tagged CTD; lane 12, D6 before concentrating; lane 13, concentrated D6. Lanes 3, 7, and 14, filtrate; lanes 4, 8 and 15, membrane washes.

Table 6.1 Molecular mass (M) estimates of protein species from GPC analysis. The proteins were loaded either individually or as a mix as described (section 6.2.2) onto a Superdex 200 column. The running buffer was 20 mM Tris-HCl, pH 7.5 with or without 150 mM NaCl. The M values have been estimated from a standard curve prepared in the appropriate running buffer.

Protein	Expected M (kDa)	M estimate (kDa) 150 mM NaCl	M estimate (kDa) No NaCl
LplA	32.6	33	39.9
CTD	12.6	11.8	22.8
E2lipD	11.3	10.4	21.1
LplA-CTD complex	44.6	47.1	-
LplA-E2lipD complex	43.9	39.4	-
E2lipD-CTD complex	23.9	-	-
Ternary complex	56.5	-	51

6.2.2.1 Activity of the ternary complex

Although it has been shown before that all the three proteins, LplA, CTD and E2lipD, are active in a gel shift assay, the ternary complex eluted from GPC (Fig. 6.6 C, lane 6) was also tested for activity to test its functional relevance. A lipoylation assay with the concentrated fraction D5 was set up as previously described in section 3.2.15. A clear shift in the activity assay confirmed that the proteins did not lose activity during the GPC analysis and that the purified ternary complex carries out lipoylation (Fig. 6.7).

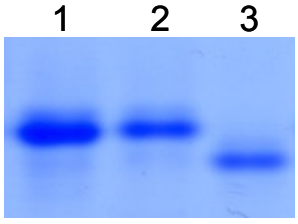


Fig. 6.7 Activity of the ternary complex isolated from GPC. Lane 1, non-lipoylated E2lipD; lane 2, reaction mix without ATP; lane 3, reaction mix with ATP.

6.2.3 Dynamic light scattering of LplA and CTD

After establishing the physical association of LplA and CTD, the proteins were analysed using dynamic light scattering (DLS) to determine their molecular sizes. In DLS, a laser beam is passed through a protein solution and any particles present in this solution observably scatter the light in a time- and intensity-dependent fashion. These scattering profiles follow the principle of Brownian motion of the particles *i.e.* their random movement. The velocity of the particles through a solution is expressed as the diffusion coefficient (D). Based on the value of D, the hydrodynamic diameter of a particle can be calculated by applying the Stokes Einstein equation:

$$d(H) = kT / 3 \pi \eta D$$

where
d(H) = hydrodynamic diameter
D = translational diffusion coefficient
k = Boltzmann's constant
T = absolute temperature
η = viscosity of the solution

The software supplied with the Malvern instrument applies the Stokes Einstein equation to calculate the hydrodynamic diameter. As this DLS value refers to how a particle diffuses within a fluid *i.e.* hydrodynamic diameter, the diameter obtained by this technique represents the diameter of a sphere that has the same diffusion coefficient as the particle. This assumption is applied to derive the M_r of protein molecules.

For consistency and reliability, the samples were subjected to three DLS runs. Samples of purified His-LplA (~ 1 mg/ml) and His-CTD (~ 1 mg/ml) in 20 mM Tris-HCl, pH 8.0, 0.2 M NaCl were prepared for DLS analysis by centrifugation at 13,000 x g for 30 min at 20°C (section 3.3.3). This step was necessary to remove large suspended particles or aggregates. DLS measurements of His-LplA at 25°C gave a molecular mass estimate of 35 kDa (± 1 SE). This value is very close to the expected 33 kDa calculated from the amino acid sequence. Measurements of His-CTD gave a molecular mass estimate of 46 kDa (± 3.2

SE). This estimate was approximately 4 times larger than the expected 13 kDa. The chromatograms of these measurements indicated polydispersity in the distribution of molecular species and the size distribution varied indicating different oligomerization states of His-CTD under the experimental conditions. In order to rule out a role of the His-tag in this process, DLS measurements of ntCTD (1 mg/ml) were carried out. The samples were prepared by the method described previously (section 3.2.11). DLS measurements of ntCTD gave an estimated 79 kDa (± 7.5 SE). This value exceeded the expected size (10 kDa) by approximately 7 times. A possible reason for such a difference in the expected and calculated molecular mass could be oligomeric self-association of ntCTD in solution. Further efforts to prevent potential protein self-association were made and a common detergent, Tween-20, was included in the protein samples. Analysis of ntCTD (1 mg/ml, ≥ 95% purity) in the presence of 0.5% (v/v) Tween-20 gave 50 kDa (± 0.5 SE). This molecular mass estimate was approximately 5 times larger than the calculated 10 kDa. Therefore, the results suggested that both His-CTD and ntCTD have a strong tendency for oligomeric self-association.

Further, the His-LplA-ntCTD complex from the pull-down assay was analysed using DLS, which gave M = 39 kDa (± 3.2 SE). The expected molecular mass for a His-LplA-ntCTD complex in which LplA and CTD are present in a 1:1 ratio is approximately 44 kDa. This result is in agreement with GPC data which also gave a size estimate of a LplA-CTD complex in a 1:1 ratio (Table 6.1). DLS analysis of mixtures of His-LplA : ntCTD in a molar ratio of 1:8 and 1:4 gave 52 kDa (± 0.6 SE) and 54 kDa (± 1.7 SE), respectively. Therefore, it appeared that His-LplA-ntCTD complex samples containing excess CTD show a molecular mass further away from the expected mass of the complex (~ 44 kDa). An explanation could be the excess of CTD, which may be forming oligomers of itself, which causes a polydispersity in the solution and leading to the skew in the results. Repeating the experiments in the presence of 0.5% (v/v) Tween-20 yielded similar results. Therefore, if oligomerization is taking place, Tween-20 is unable to break these interactions.

On the other hand, the sample from the pull-down assay gave a molecular mass much closer to the estimated value. Hence, the presence of His-LplA may be preventing the self-association of ntCTD. In other words, the His-LplA-ntCTD complex may be energetically a more favourable species than oligomers of CTD. Relatively small discrepancies in the molecular mass values using DLS can be attributed to the fact that DLS is a low resolution technique. Polydispersity and presence of large aggregates or particles do not allow for an accurate molecular mass estimation. The reason for this is that the basic distribution obtained from a DLS measurement is intensity and all other distributions (number, volume) are generated from this. It cannot be excluded that a His-LplA-ntCTD complex was present in all the different mixtures tested, but the presence of ntCTD oligomers may have masked the signal from the HisLplA-ntCTD complex.

6.2.4 Surface plasmon resonance analysis of LplA, CTD and E2lipD interaction

After establishing the association between the LplA, CTD and E2lipD, it was decided to employ surface plasmon resonance (SPR) to study their binding affinity. SPR is a sensitive technique, routinely used to quantitatively study protein-protein interactions. It exploits an optical phenomenon using a refractive index sensor. If a beam of light excites mobile electrons on the surface of a metal film, it will cause oscillations. These mobile electrons are also called plasma and hence these oscillating plasma waves are called surface plasmons. When the incident light and surface plasmon wavelength match, the electrons resonate, hence the term surface plasmon resonance. This event uses energy which causes a decrease in the angle of the reflected light. This angle is called surface plasmon angle (Θ_{spr}). Any change on the side of the metal which is not in direct contact with the light, will have an effect on Θ_{spr}. If the metal surface is immersed in an aqueous buffer, it will have a lower refractive index than when protein is bound to it. Therefore, the binding of protein causes a detectable decrease in Θ_{spr} and this signal is presented as response units (RU) (Fig. 6.8). RU is an arbitrary unit chosen so that 1 RU corresponds to a shift in angle of about 10^{-4} degrees which equals a change in refractive index of about 10^{-6}. From this, as a guideline for proteins on a CM5 sensor chip, 1 RU corresponds

to a change in surface concentration of approximately 1 pg/mm^2 (Biacore® Concentration Analysis Handbook, 2001; O'Shannessy *et al.*, 1993). Hence, RU values are proportional to the amount of protein bound to the metal surface. RU readings are presented graphically in a sensorgram which displays RU versus time (Biacore® Basics, Edition 2003).

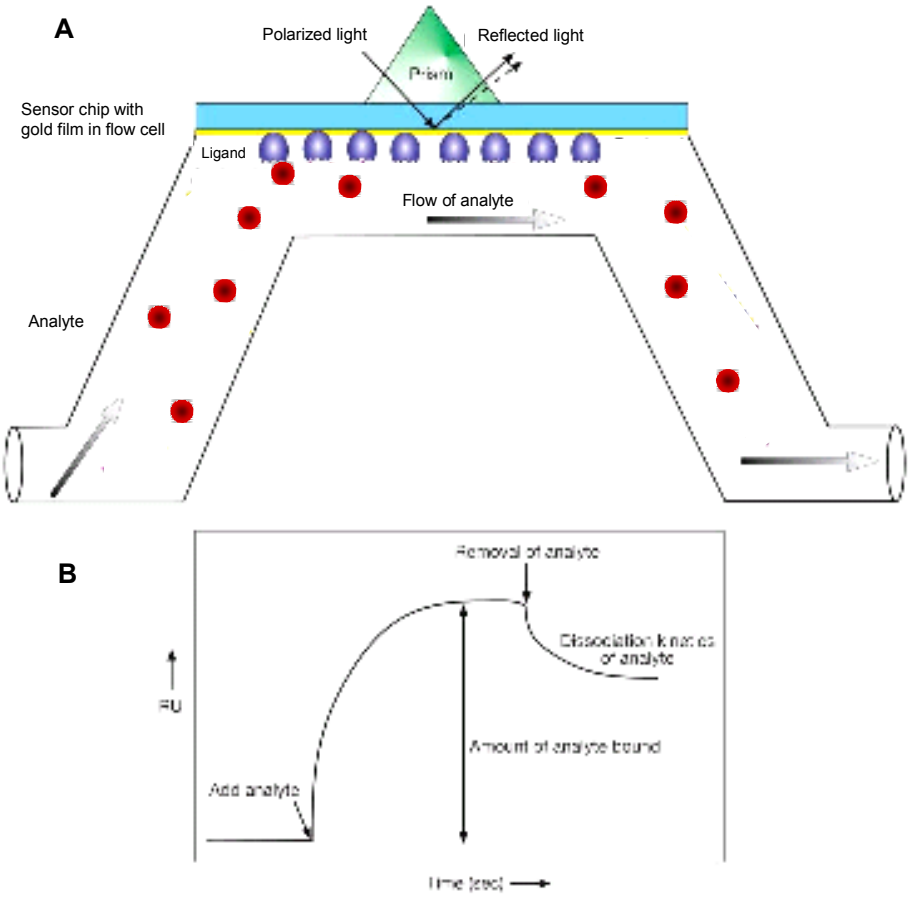


Fig. 6.8 Surface plasmon resonance (SPR). **A**: Example of a basic SPR set-up. The ligand is immobilized on the chip surface. Binding of the analyte to the ligand alters the angle of reflected light (dashed arrow) and this value can be presented in response units (RU) versus time. **B**: Example of a sensorgram showing the binding of the analyte to the ligand and the kinetics of binding and dissociation (adapted from Varki *et al.*, 2009).

6.2.4.1 SPR analysis of ntLplA and ntCTD interaction

Initial interaction studies of His-LplA and ntCTD were carried out using a CM5 chip, which had penta-His antibody immobilized on its surface (section 3.3.2). The antibody surface captured His-tagged LplA, creating a His-LplA surface. Next, ntCTD was passed over this chip in HBS-EP buffer, pH 7.4, and the spe-

cific binding response was measured. The sensorgram of His-LplA and ntCTD indicated an interaction between the two proteins (Fig. 6.9). This interaction was via His-LplA as ntCTD was not recognized by the penta-His antibody. As $MgCl_2$ is required for enzyme function ([Online] available at http://www.uniprot.org/uniprot/Q9HKT1, accessed December 15th, 2008), 5 mM $MgCl_2$ were added to the HBS-EP buffer, pH 7.4, and the experiment was repeated. In the presence of 5 mM $MgCl_2$ a higher RU was measured which indicated that $MgCl_2$ may stabilize the interaction of the His-LplA ligand with the CTD analyte. Therefore, 5 mM $MgCl_2$ was included in all subsequent experiments.

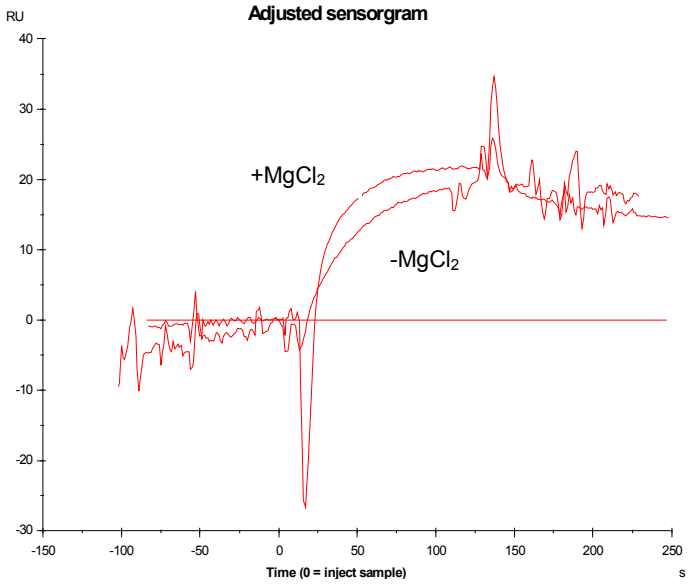


Fig. 6.9 Binding of ntCTD to His-LplA in the presence and absence of $MgCl_2$. RU, Response units; His-LplA was immobilized on a CM5 chip via penta-His antibody on the chip surface. ntCTD was passed over this chip and the binding response was recorded. This experiment was conducted in the absence and presence of 5 mM $MgCl_2$.

To characterize the binding affinity between the LplA and CTD, the non-tagged proteins were used to rule out any cross interaction of the His-tags. Both, binding kinetics (Fig 6.10) and steady state analysis (Fig. 6.11) were employed.

Kinetic analysis gives information on how fast the binding between the ligand and analyte is. This is expressed by k_a, the association constant, and is defined as the number of ligand-analyte complexes formed per unit time at unit concen-

tration of ligand and analyte (standard units $M^{-1}s^{-1}$). The typical range of k_a values is 10^{-3} to 10^7 $M^{-1}s^{-1}$. The stability of the complex in the kinetic experiment is described by the dissociation constant k_d, which is defined as the number of ligand-analyte complexes that dissociate per unit time (standard units s^{-1}). The typical range of k_d values is 10^{-1} to 10^{-6} s^{-1}. Measuring the kinetics of binding of protein interactions at different analyte concentrations can be employed to determine the equilibrium binding affinity (K_D), which equals k_d/k_a. If the kinetic curves can be fitted, K_D can be estimated without reaching steady state during the experiments.

On the other hand, steady state analysis defines how strong an interaction is *i.e.* the affinity of the interaction at equilibrium. To determine the equilibrium binding affinity constant (K_D) at steady state, data points towards the end of analyte injection are taken for different analyte concentrations. In a secondary plot, the obtained response measured at those points, referred to as R_{eq}, is plotted against the analyte concentration. From this plot R_{max} can be determined, which gives the maximal signal in RU dependent on the number of binding places. The concentration at which R_{max} is $R_{max}/2$ equals the affinity constant K_D. The typical range of K_D is $1x10^{-5}$ to $1x10^{-12}$ M. If the steady state is not reached during the injection any K_D determined this way will be an underestimate of the true value.

The interaction of ntCTD with ntLplA can be assumed to follow pseudo first-order kinetics since the concentration of ntCTD is constant throughout the injection:

$$\text{ntLplA} + \text{ntCTD} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \text{ntLplA-ntCTD}$$

where k_a is the association constant and k_d the dissociation constant.
ntLplA was immobilized onto a CM5 chip as described in section 3.3.2 and a number of different ntCTD concentrations were passed over the ntLplA surface (Fig. 6.10, Table 6.2). The response was observed to be increasing with increasing concentration of ntCTD. For analysis, the binding data was fitted to a bivalent model (Table 6.2) which gave two K_D values: K_{D1} = $1.46x10^{-4}$ M and K_{D2} = 0.7 M. Very often the bivalent analyte binding model fit is a result of slow dissociation of the ligand-analyte complex caused by multiple bond interactions.

Therefore, although the bivalent fit gives empirical values for relative rates and binding strengths, these values are not the physical constants (Biacore® kinetics and affinity analysis handbook, 2006).

The value of K_{D1} indicates a weak interaction, whereas K_{D2} is a very weak affinity constant indicating insignificant interaction between ntLplA and ntCTD. Nonetheless, the results of GPC, pull-down assay and DLS presented above showed that LplA and CTD interact specifically in a 1:1 ratio.

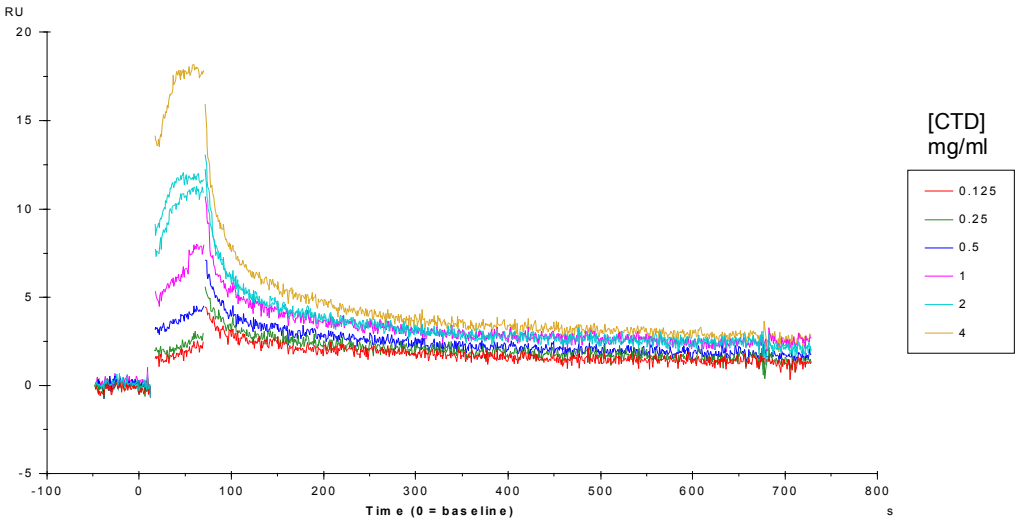


Fig. 6.10 Binding kinetics of ntCTD to ntLplA. ntLplA was covalently immobilized on a CM5 chip. The running buffer was HBS-EP buffer, pH 7.4, 5 mM $MgCl_2$. ntCTD (0.125 – 4 mg/ml) was passed over the surface for 60 s at 30 µl/min. Between different ntCTD concentrations, the ntLplA surface was regenerated with 10 mM NaOH, for 30 s.

Table 6.2 Summary of the kinetic data for binding of ntCTD to immobilized ntLplA (Fig. 6.10). k_t is the mass transport parameter, R_{max} has been fitted locally. The data fitted a bivalent model using the BIAevaluation software (Biacore®) which gave the following parameters:
k_{a1} = 170.5 $M^{-1}s^{-1}$, **k_{d1}** = 0.025 s^{-1}, **K_{D1}** = 1.46x10^{-4} M, **K_{D2}** = 0.7 M, **Chi²** = 0.147 RU^2

Cycle	[CTD] mg/ml	CTD (M)	R_{max} (RU)	k_t (RU/Ms)	RI (RU)
1	4	3.899x10^{-4}	17.98	3.217x10^{8}	3.252
2	2	1.949x10^{-4}	15.98	3.217x10^{8}	1.522
3	1	9.747x10^{-5}	15.26	3.217x10^{8}	0.3805
4	0.5	4.873x10^{-5}	14.84	3.217x10^{8}	0.002825
5	0.25	2.437x10^{-5}	17.95	3.217x10^{8}	0.004994
6	0.125	1.218x10^{-5}	23.59	3.217x10^{8}	0.001298
7	2	1.949x10^{-4}	15.5	3.217x10^{8}	0.6095

The K_D value was also determined by employing the steady state analysis (Fig. 6.11). The secondary plot from this steady state analysis (Fig. 6.12) gave a K_D value of $6.3x10^{-4}$ M. Therefore, both kinetic and steady state analyses gave K_D values indicating a weak, yet specific, interaction.

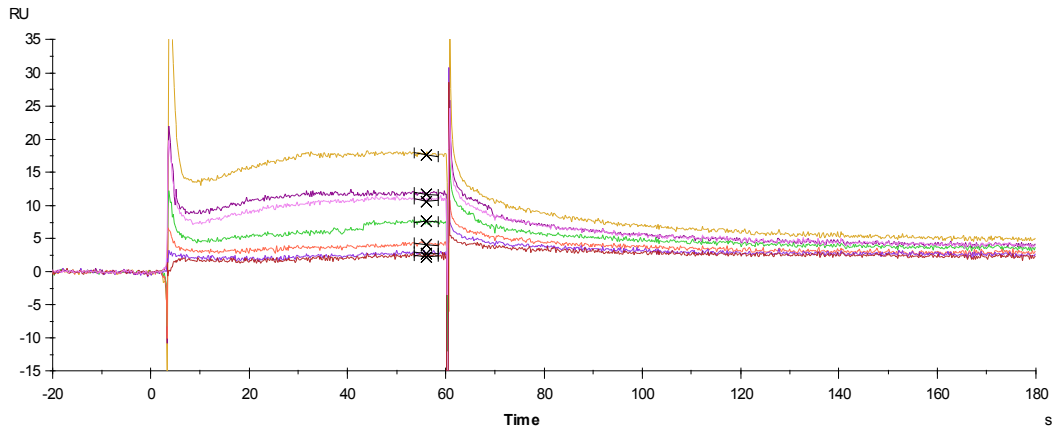


Fig. 6.11 Steady state analysis of ntCTD binding to immobilized ntLplA. The data points were taken towards the end of the injection (indicated by crosses) under the assumption that a steady state between association and dissociation exists. The analysis was done using the BIAevaluation software (Biacore®). The calculated parameters were: K_D = 6.307 x 10^{-4} M, R_{max} = 42.46 RU, Chi^2 = 0.187 RU^2

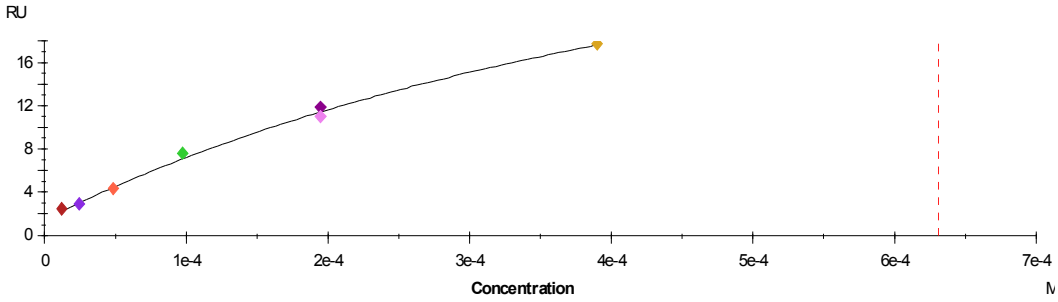


Fig. 6.12 Determination of K_D from a secondary plot of R_{eq} (units RU) versus ntCTD concentration (M). The values of R_{eq} were obtained from steady state analysis (Fig. 6.11).

Analysis of the kinetic binding (Fig 6.10, Table 6.2) and steady state (Fig. 6.11 and 6.12) data gave different K_D values, which indicated that the numerical K_D values are not very reliable. These results would require further validation. One obstacle during the kinetics experiment was the difficulty to saturate the binding of ntCTD to ntLplA. To overcome this problem, the inverse approach

was applied by immobilising ntCTD. In theory, this should give similar results as the previous approach but it seemed that ntCTD, once immobilised, could not interact with ntLplA anymore. This may be due to the unavailability of the LplA binding site of CTD in the immobilised state and/or the proteins incompatibility with the sodium acetate buffer, pH 5.0, used during the immobilization process. Moreover, kinetic analyses can be limited by mass transport effects (Nguyen *et al.*, 2007). Mass transport is solely dependent on diffusion and describes the movement of molecules from solution to a surface. It can slow down association by limiting supply of analyte to the surface-bound ligand as well as dissociation of ligand-analyte complexes because the analyte is not sufficiently washed off the surface, allowing rebinding. If data are heavily mass transport limited the k_a and k_d will be underestimated. Therefore, if the protein-protein interaction is much faster than the diffusion of CTD to the LplA surface, the observed binding will be limited entirely by the mass transport process.

6.2.4.2 SPR analysis of the E2lipD, LplA and CTD ternary complex

Previously, the pull-down experiments indicated the interaction between E2lipD with LplA and CTD. SPR experiments were conducted to further investigate the interaction of all the three components. A His-E2lipD surface was created by capturing the protein with penta-His antibody immobilized on a CM5 chip surface. After His-E2lipD had bound to the chip surface, ntLplA was passed over it. ntLplA interacted with E2lipD yielding a specific response of approximately 250 RU. ntLplA (100 µg/ml) alone yielded the same response as a mixture of equal concentrations of ntLplA and ntCTD (100 µg/ml). Interestingly, changing the concentration of LplA:CTD from 1:1 to 1:10, while keeping the LplA concentration constant, decreased the initial response to about 225 RU (Fig. 6.13). To establish if this drop in RU was due to the increase in ntCTD, the experiment was repeated with increasing concentration of CTD, keeping ntLplA constant (100 µg/ml). A drop in the RU as a function of increasing ntCTD concentration was observed indicating that the presence of ntCTD may not allow LplA to interact with E2lipD (Fig. 6.14). The GPC data have previously shown the role of NaCl in breaking up a ternary complex into two separate complexes of LplA-CTD and LplA-E2lipD. During SPR experiments, the proteins were prepared in HBS-EP buffer, pH 7.4, which contained 150 mM NaCl. Therefore, a prefer-

ence for CTD to bind to LplA is not unexpected. As LplA is still able to bind to E2lipD in the GPC experiments, the observed decrease in RU in the presence of ntCTD may indicate that the CTD-LplA interaction is more favourable in comparison to the LplA-E2lipD interaction or that CTD obstructs the binding site for E2lipD under the experimental conditions.

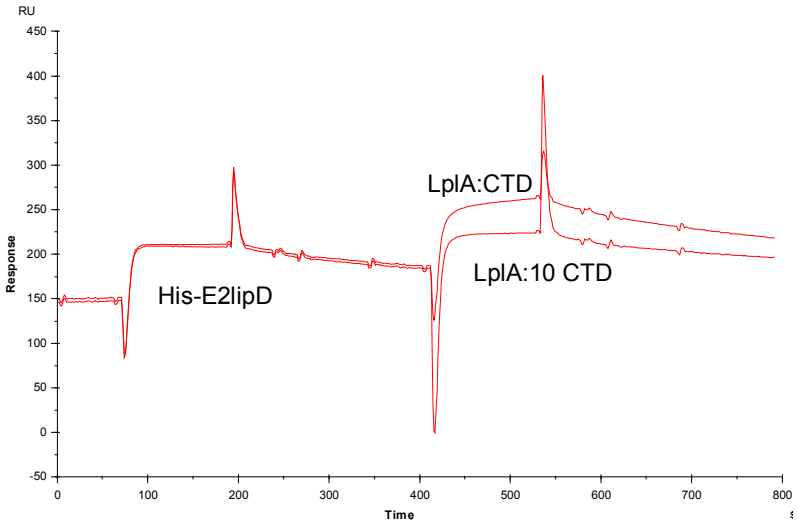


Fig 6.13 Interaction of His-E2lipD with ntLplA and ntCTD. Analytes were ntLplA and ntCTD. RU, response units. His-E2lipD (100 µg/ml) was immobilized on a CM5 chip surface via penta-His antibody at 210 RUs. ntLplA (100 µg/ml) was mixed with ntCTD (100 µg/ml) in a 1:1 and 1:10 ratio before passing it over the His-E2lipD surface.

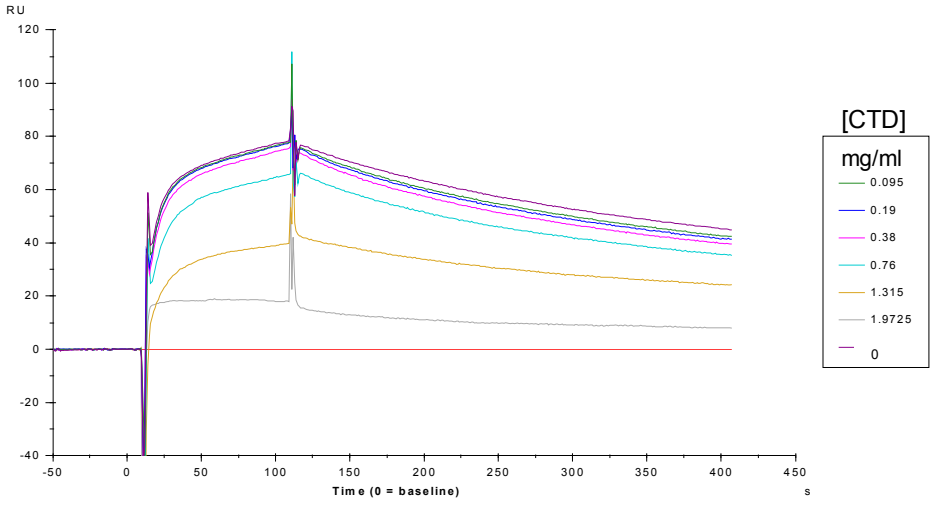


Fig. 6.14 Competition of ntLplA and ntCTD to bind to His-tagged E2lipD. The concentration of ntLplA (100 µg/ml) was kept constant while increasing the concentration of ntCTD.

Although LplA has been suggested to be responsible for converting LA into lipoyl-AMP (Kim *et al*., 2005), the results have proven that CTD is still required for the transfer of this intermediate to E2lipD. Therefore, a critical role of CTD, as seen from our preliminary SPR studies, in the interaction between E2lipD and LplA is understandable. Due to the problems with the immobilisation of CTD, it was decided not to pursue the work using SPR beyond this point. Moreover, to obtain more reliable k_a, k_d and K_D values, the experimental set up requires optimization. This includes the potential mass transport effect, as discussed above, which can be controlled by experimental conditions like increasing the flow rate (Nguyen *et al.*, 2007). The scope of this project, time and access constraints to the Biacore® instrument (Vernalis, Cambridge, UK) did not permit further optimization of the experimental conditions. Due to the ambiguities in the k_a, k_d and K_D estimates it was not possible to determine the stoichiometry of the ntLplA-ntCTD complex from these data. However, the SPR data agree qualitatively and validate the results from the pull-down and GPC experiments in that LplA and CTD interact specifically.

6.2.5 Isothermal titration calorimetry (ITC) of His-LplA and His-CTD interaction

ITC is another quantitative technique that can be used to directly measure enthalpy changes (ΔH) involved in ligand binding and/or complex formation, binding affinity and binding stoichiometry. This technique has an advantage over SPR as the interactions are analyzed in the solution phase. SPR can be used to determine binding affinity and stoichiometry, under the assumption that the proteins themselves or their interactions are not altered by the immobilization process and/or steric hindrance. If the proteins are not affected, SPR and ITC data are in good agreement (Nguyen *et al.*, 2007). During the ITC experiment, one of the molecules under study is kept in what is called the bomb (reaction vessel) and is referred to as titrate. The interacting molecule (titrant) is injected into the bomb (reaction vessel) where the two molecules can interact and the change of heat within the bomb is measured. Raw ITC data show the change in energy (µcal/sec) versus time. Each injection is indicated by a peak, where the peak intensity is proportional to the change in energy. This change in energy can be positive or negative, depending on the type of reaction. In simple

terms, negative ΔH values indicate that the interaction of the molecules released energy *i.e*. it is an exothermic reaction. Positive ΔH values show the requirement for energy to allow the interaction to occur. This is referred to as an endothermic reaction. In the beginning of a typical ITC experiment, ΔH is expected to be larger as the titrant can interact with any molecule in the bomb. Over time, this response is expected to decrease, levelling towards ΔH = 0, as the number of unbound molecules in the bomb is decreasing and all the interacting molecules are saturated with binding partners.

For ITC analysis His-LplA and His-CTD were dialysed into 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, the same buffer used in the gel shift assay. His-LplA (33 μM) was kept in the bomb (titrate) whereas His-CTD (2 μM) was titrated at different temperatures. At 25°C the data were within the exothermic region (Fig. 6.15, A). However, these responses were small, no saturation of interaction was observed and large endothermic reactions were observed towards the end of the ITC experiment. As LplA and CTD are proteins from a thermophilic organism, the temperature during the ITC experiment was increased. Although running the experiment at 40°C prevented an endothermic reaction, the interaction did not reach saturation (Fig. 6.15, B). It was not possible to fit the data to a single site, multiple site or sequential binding model using the Microcal Origin Version 5.0. software.

The presence of endothermic and exothermic reactions may be indicative of multiple interactions and not just the association of LplA and CTD. As CTD has previously shown the tendency to self-associate/aggregate (section 6.2.3), it was decided to reduce its concentration. His-CTD (0.6 μM) was kept in the bomb and His-LplA was titrated (32.6 μM) in 5 μl injections at 25°C (Fig. 6.16, A). Under these conditions, still both endothermic and exothermic reactions were observed. Next the molar ratio between His-LplA and His-CTD were increased to 1:20 for LplA to CTD. Under those conditions the raw ITC data indicated an exothermic trend (Fig. 6.16, B). Nonetheless, fitting these data to binding models proved unsuccessful.

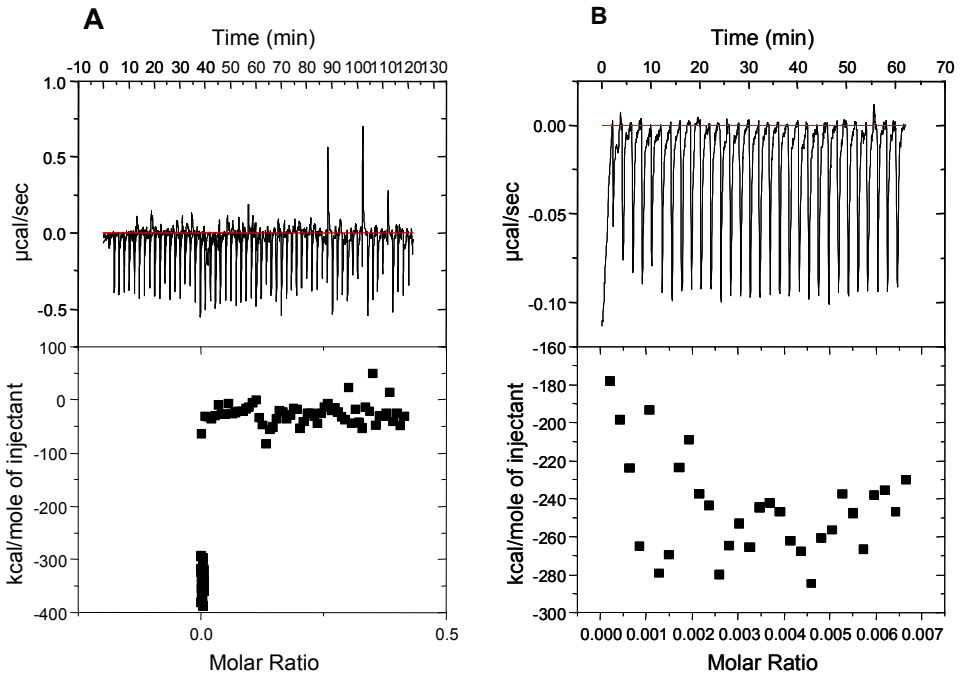


Fig. 6.15 ITC results titrating His-CTD (2 µM) into His-LplA (33 µM). The top panel shows the raw ITC data whereas the bottom panel shows the enthalpogram (ΔH data). Each experiment had 30 injections, 5 µl each. **A**: ITC at 25°C. **B:** ITC at 40°C.

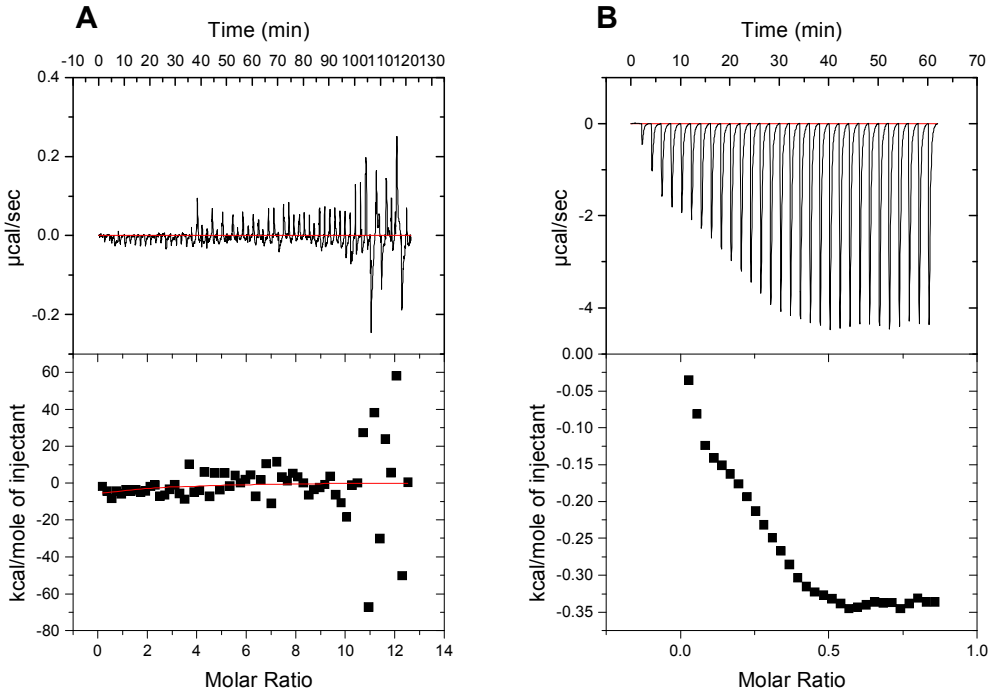


Fig. 6.16 ITC experiments titrating His-LplA into His-CTD at 25°C. The top panel shows the raw ITC data whereas the bottom panel shows the enthalpogram (ΔH data). Each experiment had 30 injections, 5 µl each. **A**: His-LplA (32.6 µM); His-CTD (0.6 µM). **B**: His-LplA (159 µM); His-CTD (7.9 µM).

6.3 Discussion and conclusions

Understanding the protein-protein interactions of LplA and CTD gives information on their possible mechanism. Either both proteins associate and catalyse the lipoylation reaction, or CTD only interacts with the LplA when a transfer of lipoyl-AMP to E2lipD is required. The latter is possible, as LplA has been shown to convert LA into lipoyl-AMP in the crystal structure (Kim *et al.*, 2005). On the other hand, the structural information of other known LplAs places the active site sandwiched between the N- and C-terminal domains. The sequence analyses and proposed three dimensional association of LplA and CTD anticipate a similar situation and is supported by the requirement of CTD for catalysis. These points suggest that LplA and CTD form a complex even in the absence of substrate and/or E2lipD.

The GPC experiments revealed three complexes: LplA-CTD and LplA-E2lipD in the presence of 150 mM NaCl and a ternary complex of LplA-CTD-E2lipD if NaCl was omitted. Hence, the ternary complex is weaker in nature and the interaction can be separated by ionic strength. This is expected from an enzymological point of view. The interaction with E2lipD should be specific for recognition and release of lipoyl-AMP, but not too strong to ensure a quick dissociation after lipoyl-AMP has been transferred. Pull-down experiments revealed that E2lipD is recognized by both, LplA or CTD. This would support the presence of a LplA-CTD complex, both proteins contributing to the active site and interacting with E2lipD. GPC on the other hand did not show an E2lipD-CTD complex. The presence of a LplA-E2lipD complex in the GPC experiments suggests that LplA is the main contributor in the recognition of the complex. Given the difference in size of LplA and CTD and that LplA makes up the majority of the active site, this is not surprising. Yet, the lipoylation activity assay shows that this is not enough for LplA to catalyze the reaction by itself.

GPC could not distinguish between the strength of the LplA-CTD and LplA-E2lipD complex. Although SPR studies did not give K_D constants of the interactions they confirmed specific weak interaction of LplA and CTD. DLS data complemented data on this interaction indicating the presence of a complex in a 1:1 ratio. Furthermore, the tendency of CTD to self-associate was suppressed

in the presence of LplA supporting that the LplA-CTD interaction is a stable and favourable protein-protein interaction.

ITC experiments indicated that the interaction between LplA and CTD may be affected by side reactions. Decreasing the CTD concentration in favour of LplA had a significant effect on the ITC data. The tendency of CTD to oligomerize/aggregate, as observed during DLS, is thought to may have affected the ITC measurements. It can be summarized as follows:

$$\text{LplA} + [\text{CTD}]_n \rightleftharpoons \text{LplA-CTD} + [\text{CTD}]_{n-1}$$

where LplA is the monomer, $[CTD]_n$ is the oligomeric form containing n copies of CTD monomers and LplA-CTD the complex between the two protein monomers. It is assumed that the endothermic reactions observed during ITC may correspond to the energy required to dissociate CTD oligomers into CTD monomers. The released CTD monomer can then interact with the LplA monomer and this interaction contributes to the exothermic response. In other words, the overall measured -ΔH is assumed to be the difference between the endothermic and exothermic reaction. The CTD oligomers and free CTD monomer are assumed to exist in equilibrium. The addition of LplA is thought to drive the formation of the LplA-CTD complex and the generation of CTD monomers pushing the equilibrium in favour of the CTD monomer. With increasing injections of LplA, the number of $[CTD]_n$ is thought to decrease and hence the observed -ΔH increases as more CTD monomers are available to interact with LplA without further energy input. This model may explain the increase in exothermic release with increasing LplA injections (Fig. 6.16, B). However, no information on how many CTD monomers may associate into oligomers, if they exist and how LplA effects their dissociation, is known. Under these conditions, ITC was not suitable to determine the K_D and stoichiometry of the LplA-CTD complex and therefore this technique was not pursued beyond this point.

Overall, these results suggest that LplA and CTD form a complex, which may interact with E2lipD to carry out the lipoylation reaction. This interaction may not depend on the presence of substrate or reaction intermediate, which suggests that major substrate induced conformational changes are less likely to

play a role in the lipoylation reaction. Therefore, it is possible that LplA and CTD may form a stable complex which may not dissociate after assembly, similar to other LplAs (Cronan *et al.*, 2005; Fujiwara *et al.*, 2005; Fujiwara *et al.*, 2007) which exist as one polypetide. In the following chapter, an attempt has been made to produce a fusion protein consisting of LplA and CTD and its enzyme activity has been analyzed.

CHAPTER 7

Design of a *Thermoplasma acidophilum* lipoate protein ligase and C-terminal fusion protein (LplA-CTD Fusion)

7.1 Introduction

In the previous chapters it was demonstrated that both LplA and CTD are required for lipoylation and that they physically associate to do so. With reference to the protein sequence alignments, the *Tp. acidophilum* CTD shows approximately 30% sequence homology to the C-terminal domains of other lipoate protein ligases. However, in the case of *Tp. acidophilum*, CTD is not part of the same polypeptide but is present as a separate gene directly upstream of the truncated *lpla* gene (chapter 4). Consequently, it came into question if a fusion protein of *Tp. acidophilum* LplA and CTD would be active. An answer to this question may help enhance our understanding of their activity and may provide a possible evolutionary link between the lipoate protein ligase from *Tp. acidophilum* and that from other species. As this is the first evidence for lipoylation in the Archaea, one has to be careful in making swift comparisons. It might be possible that the *Thermoplasma* LplA protein, encoded by two genes rather than one, has a different structural arrangement that requires the two domains to be present as separate polypeptides to accommodate their association. To evaluate this possibility, a fusion of both genes in the hypothetical N- to C- terminal position, as indicated by sequence alignments (chapter 4), was designed and analysed for lipoylation activity.

7.2 Results

7.2.1 Design and cloning of a LplA-CTD Fusion protein

Based on protein sequence identity of the individual LplA and CTD with the N- and C-terminal regions of other known LplAs respectively, the CTD gene was relocated downstream of the LplA gene. The predicted fusion product was used in a three-dimensional structure model comparison with *S. pneumoniae* LplA (Fig. 7.1).

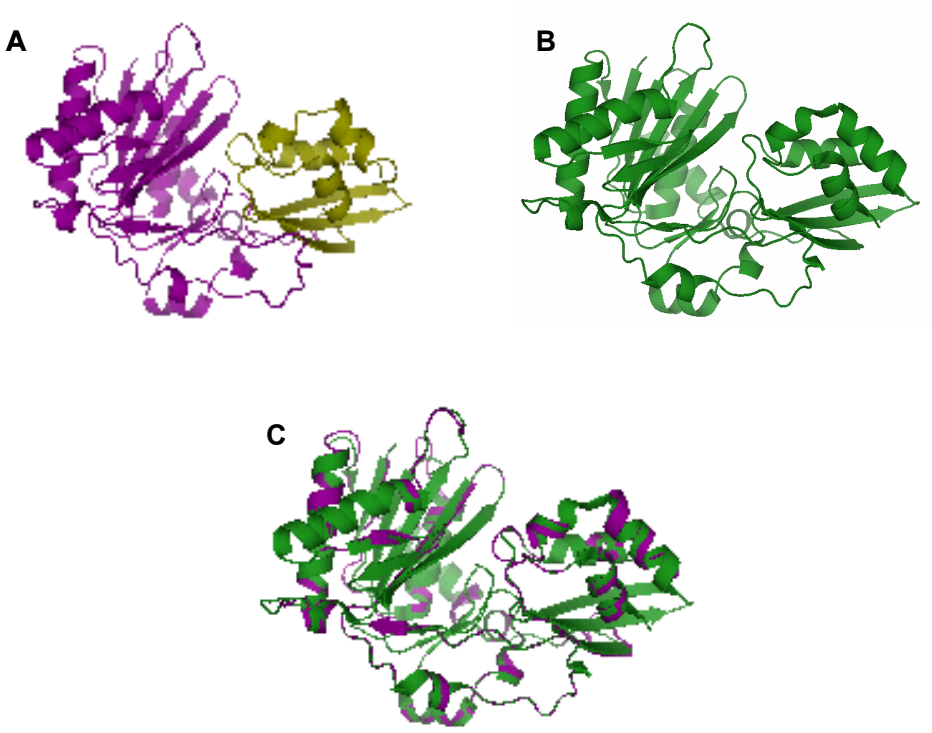


Fig. 7.1 Three-dimensional structure model comparison of the *Thermoplasma* LplA-CTD Fusion protein with *Streptococcus pneumonia* LplA. **A**: LplA-CTD Fusion model; the LplA (N-terminal domain) and CTD (C-terminal domain) are highlighted in magenta and yellow, respectively. **B**: *S. pneumoniae* crystal structure (PDB accession 1VQZ) (green). **C**: Overlay of the predicted *Thermoplasma* LplA-CTD Fusion (magenta) structure model with the *S. pneumoniae* LplA (green) structure. The 3-D coordinates of the LplA-CTD Fusion were generated from its amino acid sequence using the program Quickphyre (Bennett-Lovsey *et al.*, 2008; accessed July 10th, 2008) and from these data a 3-D model was generated using PyMOL (DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org).

In order to express the *Tp.* LplA together with its hypothetical C-terminal domain in the proposed order, a physical association between the two genes, *lpla* (*ta0514*) and *ctd* (*ta0513m*), was created using overlap extension PCR (OEPCR) (Fig. 7.2). In this design, the *lpla* stop codon and *ctd* start codon were deleted and the two genes fused without a linker. The 3` region of *lpla* was fused with the 5` region of *ctd* by designing the reverse primer for *lpla* and the forward primer for *ctd* as homology primers. The forward primer for *lpla* was 5'-GCTAGC*ATG*GAAGGCAGGCTTCTTTTACTAGAA-3' (start codon in italics and NheI restriction site underlined) and the reverse homology primer was

5'-GTACATCATATGTACGACCTCTTTCCTCAGGAG-3', where the black and the blue match the gene sequence coding for the 5` end of *ctd* and the 3´ end of *lpla,* respectively (Fig. 7.2). The forward homology primer for *ctd* was 5'-AGGAAAGAGGTCGTACATATGATGTACAGCAAG-3', where the black and green match the DNA sequence coding for the 3`end of *lpla* and 5´end of *ctd,* respectively (Fig. 7.2), and the reverse primer was 5'-CTCGAG**TCA**GATCACCCTCAAAGCCTGAAT-3' (stop codon in bold and XhoI restriction site underlined). The restriction sites were introduced for ligation of the PCR product into the pET28a vector and for the addition of an N-terminal His-tag. In the first PCR reaction *lpla* and *ctd* were amplified individually. The use of homology primers allowed overlap and hence amplification of the two DNA sequences as a whole in a defined manner in a second PCR reaction (Fig 7.2).

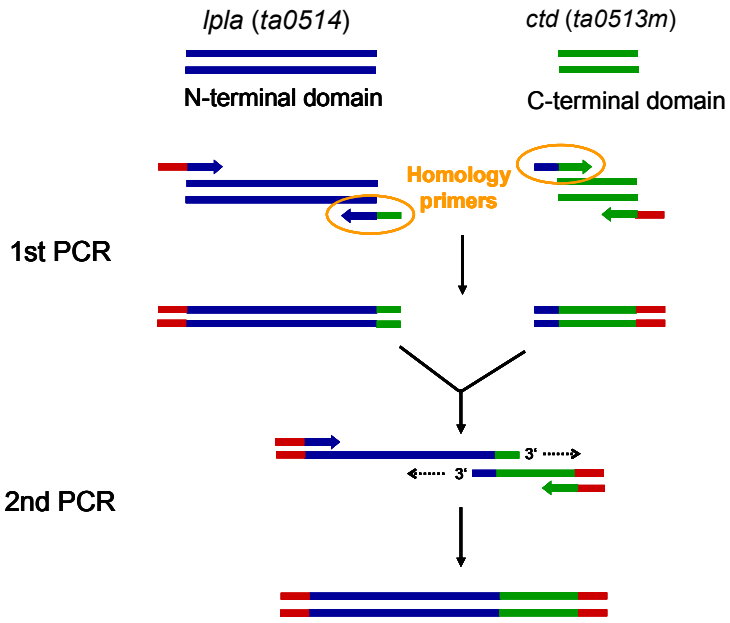


Fig. 7.2 Diagram of the OEPCR reaction. The gene sequences coding for *Tp.* LplA and the CTD were first amplified individually. The PCR conditions were the same for both genes with denaturation at 98°C for 30 s followed by 30 cycles with denaturation at 98°C for 10 s, annealing at 64°C for 30 s and extension at 72°C for 30 s. Final extension was allowed to take place at 72°C for 10 min. In the second PCR, the purified PCR products (section 3.1.3), the forward primer for *lpla* and the reverse primer for *ctd* were added together. Initial denaturation was at 98°C for 30 s and was followed by four cycles of denaturation at 98°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 2 min to allow overlap of the two PCR products at DNA regions created by the homology primers in the first PCR reaction. The reaction was completed with 25 subsequent cycles with denaturation at 98°C for 10 s, annealing at 67°C for 30 s and extension at 72°C for 30 s. Final extension was done at 72°C for 10 min.

OEPCR successfully produced the desired *lpla-ctd fusion* gene (Fig. 7.3). The *lpla-ctd fusion* gene was cloned into pGEM-T Easy vector and sent for sequencing. A clone that had the correct sequence was ligated into the pET28a vector.

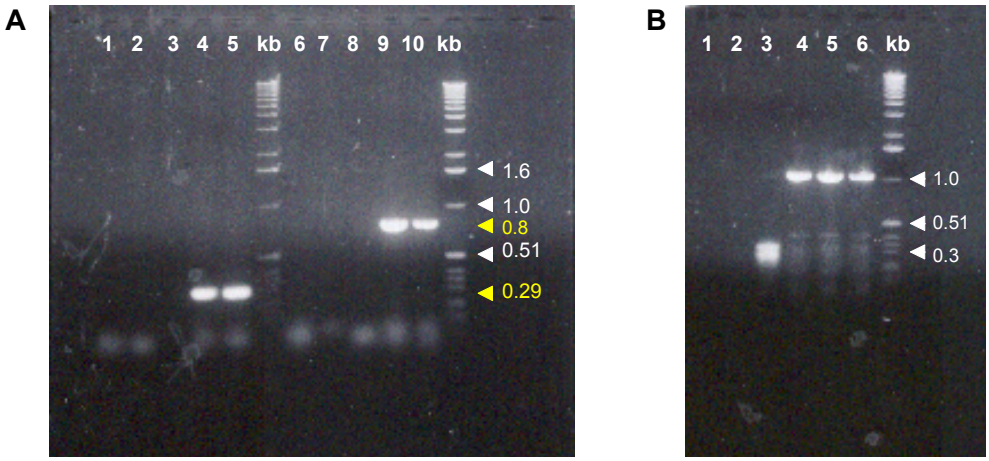


Fig. 7.3 PCR products of the individual PCRs and OEPCR (1% agarose gels). Marker sizes are in kilo base pairs. **A**: Individual amplification of *ctd* (lanes 1 – 5) and *lpla* (6 – 10). Lanes 1 – 3 and 6 – 8 are the controls in which DNA, the reverse primer and forward primer were omitted, respectively. Lanes 4, 5, 9 and 10 are the complete PCR mixtures. The amount of template DNA in reaction 4, 5, 9 and 10 was the same, but the reaction mixtures in lane 5 and 10 contained DMSO. Both PCR products are of the correct size (yellow). **B**: OEPCR. Using the purified PCR products from A as template in a second PCR, a PCR product of the correct size (1080 bp) was obtained. Lane 1, 2 and 3 are the controls in which the DNA, reverse primer and forward primer were omitted respectively. The control in lane 3 gave some PCR product corresponding in size to the *ctd*. Lane 4, 5, and 6 are the complete PCR mixtures. The reaction mixtures in lane 5 contained twice the amount of template DNA. Lane 6, reaction mixture included DMSO.

7.2.2 Expression and purification of the LplA-CTD Fusion protein

The *lpla-ctd-fusion*-pET28a plasmid was transformed into BL21(DE3) *E. coli* cells. Cell cultures were grown in OvernightExpress autoinduction medium (Novagen) overnight at 37°C with shaking. Cells were harvested and sonicated in lysis buffer (20 mM Tris-HCl, pH 8.0, 0.3 M NaCl, 20 mM imidazole) and the cell lysate was centrifuged at 75,000 x g for 20 min at 4°C. The supernatant was subjected to Ni^{2+} affinity chromatography on a HisTrap FF column applying a linear gradient of imidazole (0.02 – 0.5 M). The fractions containing the LplA-CTD Fusion were diluted 10-fold in lysis buffer and the previous purification step was repeated. The purified LplA-CTD Fusion had a relative electrophoretic mo-

bility of approximately 45 kDa on SDS-PAGE gel, in agreement with the expected 42 kDa calculated from the amino acid sequence (Fig. 7.4).

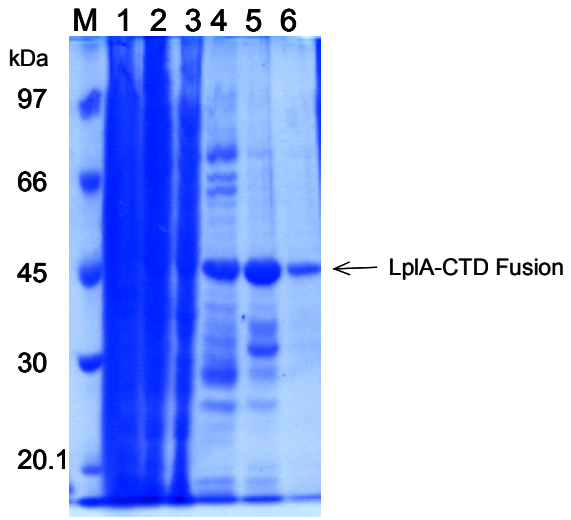


Fig. 7.4 SDS-PAGE gel summarizing the Ni^{2+} affinity purification of LplA-CTD Fusion on a HisTrap FF column. M, protein marker (kDa); lane 1, whole cell sample; lane 2, whole cell sample after induction with IPTG; lane 3, soluble fraction of the IPTG induced whole cell sample (lane 2); lane 4, LplA-CTD Fusion from first Ni^{2+} affinity purification; lanes 5 and 6; LplA-CTD Fusion from a second Ni^{2+} affinity purification step. LplA-CTD Fusion is indicated by the arrow.

7.2.3 Lipoylation activity of the LplA-CTD Fusion protein

A gel shift assay was set up with purified LplA-CTD Fusion as described (section 3.2.15). In the presence of the substrates LA and ATP, a clear shift in the relative mobility of E2lipD was observed, indicating that the LplA-CTD Fusion has lipoylation activity (Fig. 7.5). The amount of E2lipD being lipoylated increased with increasing the amount of LplA-CTD Fusion used in the assay.

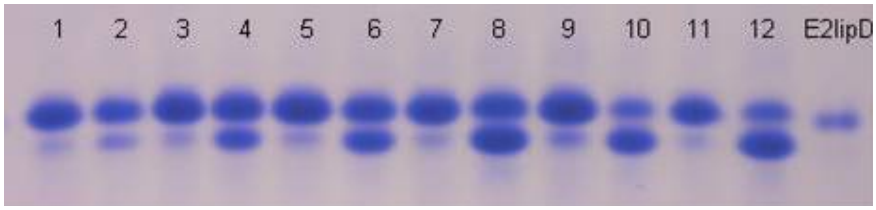


Fig. 7.5. Non-denaturing PAGE gel showing the lipoylation activity of LplA-CTD Fusion. The lipoylation reaction was set up as described (section 3.2.15). Different amounts of LplA-CTD fusion were analysed: 10 µg (lane 1, 2), 25 µg (lane 3, 4), 50 µg (lane 5, 6) µg, 100 µg (lane 7, 8), 150 µg (lane 9, 10), 200 µg (lane 11, 12). Lanes 1, 3, 5, 7, 8 and 11 did not contain ATP. Lanes 2, 4, 6, 8, 10, 12 were supplemented with ATP.

7.3 Discussion and conclusions

Using OEPCR it was possible to design, express and purify a fusion protein with LplA as the N-terminal and CTD as the C-terminal domain of the protein. Structure predictions suggested that the domains would fold as individual domains similar to the domains of the *S. pneumoniae* LplA, and it was anticipated that the domains would still fold properly after fusion. The activity assays showed that the LplA-CTD Fusion carried out lipoylation of *Thermoplasma* E2lipD. Therefore, the physical association of LplA and CTD without a linker region did not impair enzyme activity. As the absence of a linker domain between LplA and CTD did not impair enzyme activity, it suggested that the two proteins folded properly and that the arrangement of the fused LplA and CTD are likely to be the positions adopted by the individual LplA and CTD. This in turn also indicated that the interaction between the two individual proteins may be similar to the interaction of the two domains if present in one polypeptide as found in other organisms.

In the gel shift assay, 200 µg of LplA-CTD Fusion was sufficient to lipoylate about 70% of the E2lipD. The LplA and CTD have been fused without a linker, which may explain a reduced activity in comparison with the two individual proteins. The design of fusion proteins with various size linkers and testing the effect of linker size on activity could give essential information on the flexibility requirement between the LplA and CTD domains. If more time had been available, this work would have been included.

Having established that *Tp. acidophilum* has a lipoylation machinery, which resembles other lipoate protein ligases, the lack of BCOADHC activity in *Tp. acidophilum* extracts needs to be explained. The following chapter will describe evidence of *in vivo* lipoylation in *Thermoplasma* cell extracts.

CHAPTER 8
In vivo lipoylation of E2

8.1 Introduction

Recombinant expression of the *Tp. acidophilum oadhc* genes in *E. coli* has been shown to produce a functional OADHC with substrate specificity for branched-chain 2-oxoacids and, to a minor degree, pyruvate (Heath *et al.*, 2007). A recent report has also described the isolation of the BCOADHC components from *Thermoplasma* cell extracts using glycerol density gradient ultracentrifugation coupled to two-dimensional gel electrophoresis and mass spectrometry (Sun *et al*., 2007). This presence of BCOADHC components is somewhat contrary to the established fact that no *in vivo* OADHC activity in *Tp. acidophilum* cell extracts has ever been detected. As outlined in previous chapters, lipoylation is a crucial step for overall OADHC activity. The innate lipoylation machinery of *Tp. acidophilum* was the main focus in chapters 4 to 6 and the results showed that *Tp. acidophilum* has the two genes, which, if recombinantly expressed in *E. coli*, carry out lipoylation of E2lipD and whole E2 *in vitro*. The presence of LplA in cell extracts has been confirmed (Sun *et al.*, 2007, supplement) but not the presence of CTD. Therefore, lack of *in vivo* lipoylation may explain the absence of OADHC activity.

The particular lipoate protein ligase from *Tp. acidophilum*, which has been characterized in previous chapters, requires externally supplied LA as a substrate. The standard growth medium, as used by Sun *et al.* (2007) and described in section 3.1.5, was not supplemented with LA. In the proteomics analysis by Sun *et al.* (2007) it was not mentioned whether E2 was lipoylated or not. It is possible that the supply of LA induces the *Tp. acidophilum* lipoylation machinery, which in turn gives rise to an active BCOADHC. The following work is the study of the growth of *Tp. acidophilum* in the presence of LA and its effect on *in vivo* lipoylation of native E2.

8.2 Results

8.2.1 Growth of *Tp. acidophilum* cell cultures

Two *Tp. acidophilum* cell cultures were grown as described in section 3.1.5. One culture was grown in standard DSM 158 medium. To test the effect of LA on cell growth and viability, the second culture was grown in standard growth medium supplemented with 0.2 mM LA. Cell viability was checked using the microscope. The cell growth in the two cultures was measured at OD_{600nm} and growth curves were generated (Fig. 8.1). A comparison of the two growth curves showed that, despite the cell density of the standard cell culture being higher on average, both growth curves follow a similar trend. The doubling time of the standard culture is 8 h and for the LA supplemented culture 8.5 h. Performing a one-way ANOVA analysis with a p-value of 0.704 showed that this difference in doubling time is not significant. Therefore, it was concluded that LA supplement did not affect the growth of *Thermoplasma* cells cultures.

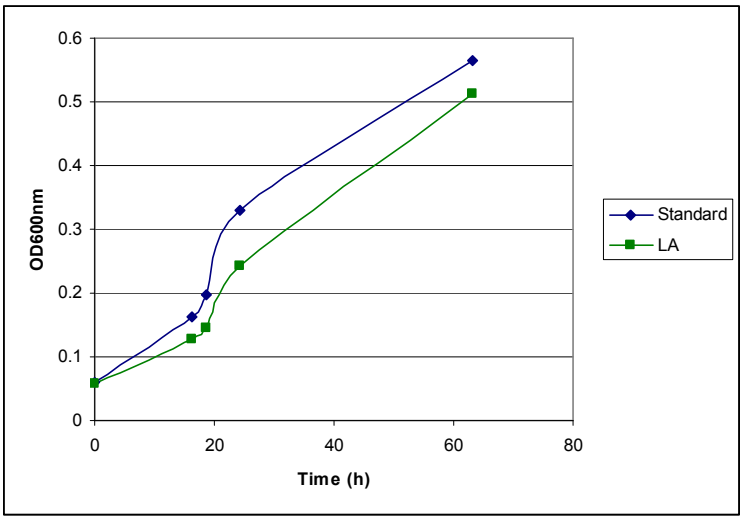


Fig. 8.1 Growth curves of *Tp. acidophilum* with and without lipoic acid (0.2 mM) supplement. 100 ml cultures were grown as described (section 3.1.5). Viability of the cells was checked under the microscope and cell density was measured at OD_{600nm}.

8.2.2 Western blot analysis of *in vivo* lipoylation

From the growth studies it was concluded that LA supplementation did not interfere with cell viability and growth. However, to assess whether the LA supple-

ment induced *in vivo* lipoylation, a Western blot of whole *Tp. acidophilum* cell extracts was carried out as described in section 3.2.17 (Fig. 8.2).

As positive controls of lipoylation, *E. coli* extracts as well as the lipoylated *Thermoplasma* E2 from *in vitro* assays were included in the Western blot analysis. *E. coli* has a PDHC and OGDHC, each containing a lipoylated E2, which run with a relative mobility of 81 kDa and 51 kDa respectively on SDS-PAGE (Perham and Thomas, 1971). The Western blot with anti-lipoic acid antibody revealed several lipoylated species in *E. coli* extract (Fig 8.2, B lane 1). The estimated sizes corresponded to species running between 83 – 61 kDa (top band, Fig 8.2, B, lane 1), 50, 36 and 20 kDa. The 50 kDa and ~ 83 kDa species could be the E2 component of the OGDHC and PDHC respectively. The broadening of the band pattern on the PAGE for the PDHC E2 component may have been a result of overloading the *E. coli* cell extract on the SDS-PAGE gel. Moreover, previous reports show that freezing and thawing of *E. coli* cultures causes E2 degradation, the main degradation product with a size of 36 – 40 kDa (Perham and Thomas, 1971). It was very likely that E2 may have been degraded as a 36 kDa and 20 kDa species were detected by the anti-lipoic acid antibody in *E. coli* cell extracts.

Lipoylated *Thermoplasma* E2 from the *in vitro* assay gave a strong signal on the Western blot (Fig. 8.2, B, lane 7) running on the SDS-PAGE gel with a relative mobility of approximately 55 kDa. In addition, bands corresponding to LplA and CTD were visible on the Coomassie stained SDS-PAGE, but not on the Western blot. This result is consistent with the previous Western blot analysis of *in vitro* lipoylation of purified whole E2 (Fig. 5.1, D). In addition to the 55 kDa species, a band around 37 kDa was present on SDS-PAGE. This band was detected by anti-lipoic acid antibody in the Western blot analysis, and maybe the results of E2 degradation similar to that observed with *E. coli* E2 protein.

The *Tp. acidophilum* cell extracts, whether grown in standard medium or on LA supplement, were similar in appearance on the SDS-PAGE gel (Fig. 8.2, A). Western blot analysis revealed a lipoylated protein species in cell cultures which had been supplement with LA for 13 days (Fig. 8.2, lane 6), whereas the cell

extract from cultures grown for 13 days on standard medium did not show any sign of lipoylation (Fig. 8.2, lane 2). The relative mobility of this lipoylated species on SDS-PAGE was approximately 55 kDa, consistent with the relative mobility of lipoylated E2. The Western blot analysis was repeated three times.

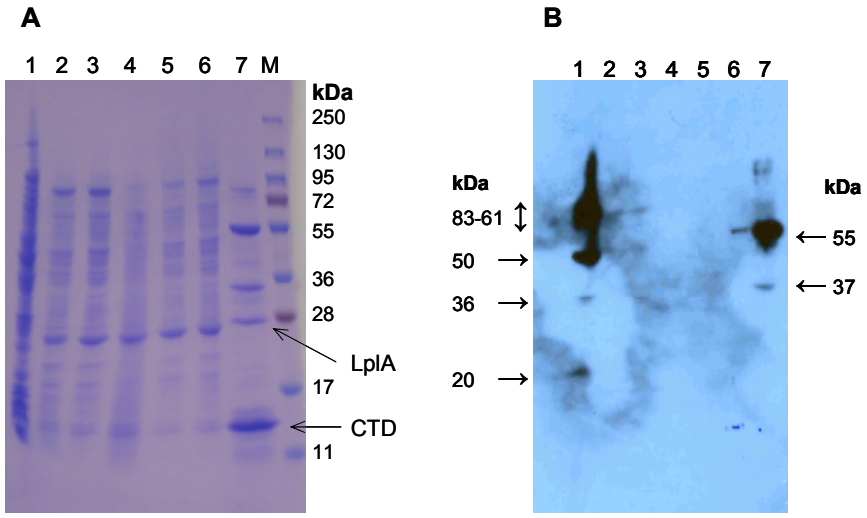


Fig. 8.2 Western blot analysis of *in vivo* lipoylation in *Tp. acidophilum*. All cell extracts were taken from cultures with an OD_{600nm} ~ 0.5. M, protein marker (kDa); Lane 1, *E. coli* whole cell extract; lane 2, *Tp. acidophilum* whole cell extract from cultures grown for 13 days in standard medium; lanes 3 – 6 *Tp. acidophilum* cell cultures grown continuously on LA since the first subculture into the supplemented medium for 6, 8, 11 and 13 days, respectively; lane 7, *in vitro* lipoylation assay of whole E2 (section 5.2.1): **A**: SDS-PAGE of the whole cell extracts and lipoylated E2 control. **B**: Western blot of A using anti-lipoic acid antibody.

8.3 Discussion and conclusions

Gene induction by the presence or absence of metabolites is a common phenomenon (Dworking and Losick, 2001; Thevelein *et al.*, 2008) and has been found to affect *oadhc* genes from *Pseudomonas putida* (Madhusudhan *et al.*, 1999). Moreover, this phenomenon has been reported in the Archaea, for example the induction of ß-glycosidase activity in *Sulfolobus solfataricus* in sucrose minimal medium without added yeast extract (Haseltine *et al.*, 1999). As previous studies showed that the BCOADHC components are present in *Thermoplasma* cultures grown on standard medium (Sun *et al.*, 2007) it was of interest to see if the *Thermoplasma* lipoylation machinery could be induced by its substrate LA. The growth studies did not show any effect of LA on cell viability or growth rate.

The Western blot analysis identified the presence of a lipoylated protein species in *Thermoplasma* cultures grown on LA for 13 days whereas the control culture grown on standard medium for 13 days did not contain any lipoylated species. The lipoylated protein species corresponded in electrophoretic mobility to recombinantly expressed and lipoylated *Thermoplasma* E2. The size of the lipoylated species, together with the limited targets of lipoylation, which only include the E2 component of OADHCs and the H-protein of the glycine cleavage system (Cronan *et al.*, 2005), led to the conclusion that it is indeed lipoylated whole E2. This is the first report of *in vivo* lipoylation in the Archaea and shows that the presence of LA is required for this to happen.

As LplA has been detected in glucose grown medium (Sun *et al*., 2007), it is interesting that the appearance of lipoylation took 13 days. This may indicate that CTD, which is required for lipoylation (chapter 5), is not constitutively expressed. The LA supplement may initiate changes at the transcriptional and/or translational level required to express the complete *Thermoplasma* lipoylation machinery. A mechanism of gene regulation at the transcriptional level has been shown for other archaea (Tang *et al.*, 1999; Haseltine *et al.*, 1999; Jenal *et al.*, 1993; Kanai *et al.*, 2006). This is an exciting new finding and holds clues to the lack of *in vivo* OADHC activity in *Tp. acidophilum*. Sun *et al.* (2007) showed that the BCOADHC components are present in *Tp. acidophilum* cell extracts, although no assessment of the levels and physiological significance in the cell had been addressed. Together with the findings presented in this chapter, it is possible that LA is required for LplA and CTD expression, which subsequently could carry out E2 lipoylation. Alternatively, only basal levels of the BCOADHC may be present in the cells and LA could be a requirement for the biosynthesis of both, the lipoylation machinery as well as the BCOADHC. To explore these possibilities, the next step is to measure BCOADHC activity in cell extracts from *Thermoplasma cultures* grown on LA. Future avenues of study include MS analysis of the lipoylated protein species from cell extracts and microarray studies to monitor the effect of LA supplementation on gene expression to complement the current data.

CHAPTER 9
Assembly of a BCOADHC from *Tp. acidophilum*

9.1 Introduction

Members of the OADHC family that have been characterized, have one thing in common: the E2 forms the catalytic and structural core of the complex (Perham 2000). In the case of PDHCs from Gram-negative bacteria and all OGDHCs and BCOADHCs, the E2 core is composed of 24 subunits with octahedral symmetry. The PDHCs of eukaryotes and Gram-positive bacteria have E2 cores made up of 60 subunits with icosahedral symmetry (Izard *et al.* 1999). Based on its substrate specificity, the *Tp. acidophilum* complex is a BCOADHC (Heath *et al.,* 2007) and an octahedral E2 core would be envisaged. Nonetheless, as no other archaeal OADHCs have been described yet, experimental data were required to confirm this assumption. This chapter describes the determination of the molecular mass of the complex proteins, the *Tp. acidophilum* E2 core and fully assembled complex and the effect of temperature on complex assembly using dynamic light scattering (DLS) and analytical ultracentrifugation (AUC).

9.2 Results

9.2.1 Optimization of NaCl concentration in DLS

The DLS measurements were carried out as described in section 3.3.3. The shown results are reproducible and were run at least in triplicate. Initial DLS experiments with purified complex components resulted in a high polydispersity index, indicating the presence of large particles and/or different oligomerization states. In such instances, DLS data are not reliable (section 6.2.3) and attempts were made to prevent the protein self-oligomerization. Presence of salt is known to prevent non-specific binding and oligomerzation by promoting preferential hydration with the aqueous environment (Hofmeister effect) (Thomas and Elcock, 2006; Baldwin, 1996). To assess the effect of salt, E1 was subjected to DLS measurement with different NaCl concentrations (Fig. 9.1).

E1 with 100 mM NaCl measured at 55°C for 10 min gave a molecular mass estimate of 168 (± 6 SE) kDa. This estimate is close to the expected molecular mass of 165 kDa, consistent with an $\alpha_2\beta_2$ conformation (Heath *et al.*, 2007). Depending on the source, E1 can exist as a homodimer (α_2) (Schreiner *et al.*, 2005; Lessard and Perham, 1994). If 100 mM NaCl were added after the heat incubation step, the M estimate was found to be 81.6 kDa. The expected molecular mass for E1 in α_2 conformation is 86 kDa, which is close to the measured estimate. Although theoretically possible, an α_2 conformation is thought to reflect a non-native oligomerization state under the given experimental conditions as the $\alpha_2\beta_2$ conformation has been found to be the only active E1 species (Heath *et al.*, 2007). Including 100 mM NaCl in the DLS measurements of E1 gave highly reproducible data, distinct from the other NaCl concentrations. Therefore, it was concluded that 100 mM NaCl ensures sample stability by preventing non-specific oligomerization. Moreover, the determined molecular mass correlated well with the expected molecular mass of the assembled E1 (Fig. 9.1).

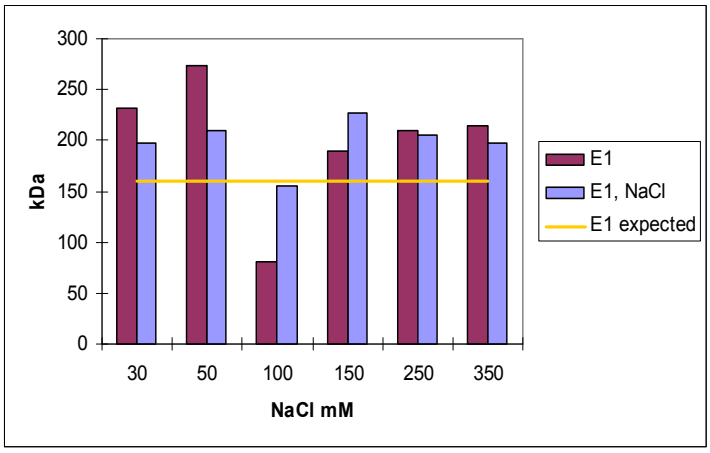


Fig. 9.1 Estimation of optimum NaCl concentration for molecular mass estimates of E1 by DLS. All experiments were measured at 25°C. Before the measurement, samples were filtered and incubated at 55°C for 10 min in the presence (blue) or absence (NaCl added after samples had cooled down) (red) of NaCl. The expected molecular mass of E1, based on $\alpha_2\beta_2$ assembly, is indicated (yellow line). 100 mM NaCl gave a molecular mass estimate closest to the expected value if the E1 sample had been heated in the presence of NaCl.

Next, the effect of 100 mM NaCl on the E2 and E3 samples was analyzed, to test if this NaCl concentration was suitable to measure their individual molecular mass. E3 is reportedly a dimer with M = 99.7 kDa (Heath *et al.*, 2007). In the presence of 100 mM NaCl, independent of its presence or absence during the incubation step, DLS measurements at 25°C gave molecular mass estimates close to the expected value (Table 9.1). DLS measurements at 55°C showed that E3 is the least affected by the presence of NaCl in the incubation step as consistent molecular mass estimates were obtained across all conditions (Table 9.1).

Table 9.1 Molecular mass estimates of E3. Prior to all DLS measurements, E3 was incubated at 55°C for 10 min in the presence or absence of NaCl. Samples that did not include 100 mM NaCl during the incubation step were supplemented with the NaCl after they had cooled down to 25°C. DLS measurements were taken at 25°C and 55°C. The molecular mass estimate is given in kDa ± SE.

	Incubation step	**25°C**	**55°C**
E3 (kDa)	No NaCl	115 ± 2.7	121 ± 0.8
	NaCl	117 ± 1.4	115 ± 1.7

DLS analysis of the E2 component was consistent with the E1 and E3 results: the presence of 100 mM NaCl was optimum to get highly reproducible molecular mass estimates which were close to the expected molecular mass of the E2 monomer (46,276 Da). Interestingly, the molecular mass estimate of E2 was affected by the presence or absence of NaCl during the heat incubation step (Fig. 9.2). If NaCl was absent from the heat incubation step and only added later, the molecular mass estimate (54,400 Da) was close to the mass of the E2 monomer. If NaCl was present during the heat incubation step, a molecular mass estimate of 1.2×10^6 Da was obtained. The mass of this species did not change during the DLS measurements and was observed throughout under the given conditions. If the *Thermoplasma* BCOADHC had a 24-mer E2 core, as predicted, the molecular mass of this core is expected to be 1128 kDa. This value is close to the mass estimate of the E2 if 100 mM NaCl were included in the heat incubation step. Therefore, this mass estimate was thought to represent the assembled E2 core of the *Thermoplasma* BCOADHC.

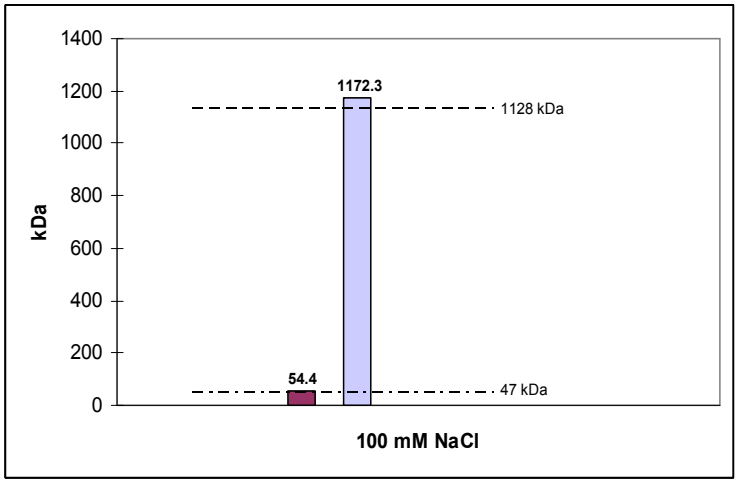


Fig. 9.2 DLS molecular mass estimates of E2. DLS measurements were done at 25°C. All proteins were incubated at 55°C for 10 min either with (blue) or without (magenta) 100 mM NaCl. 100 mM NaCl were added to samples, which had not been incubated with NaCl at 55°C, after they had cooled down. The expected molecular mass of the E2 monomer (47 kDa) and of a 24-mer core (1128 kDa) are indicated.

9.2.2 Temperature dependence of E2 core and BCOADHC assembly

Apart from the presence of NaCl, the other factor in the measurements was the temperature. In the initial observation for the E2 core, the 100 mM NaCl had to be available at a temperature of 55°C, whereas 100 mM NaCl at 25°C gave the E2 monomer molecular mass.

To investigate the effect of temperature, DLS readings of E2 supplemented with 100 mM NaCl after the heat incubation step, were performed over a range of temperatures (25°C – 60°C). The transition of the E2 monomer to a 1.0 (± 0.1) × 10^6 Da protein assembly was observed at a critical temperature of 40°C. The size was maintained up to 60°C, but aggregation was observed above this temperature. The assembled E2 core remained stable and was maintained even when the sample was cooled down to 25°C. This observed molecular mass can be explained by the assembly of a 24-mer E2 core. Indeed, this correlates with the result from analytical ultracentrifugation (Heath *et al.*, 2007).

Based on the success of using DLS in estimating the size of E2 core assembly, the technique was extended to study the whole complex assembly at 25°C and 55°C. As the E2 core is crucial for the complex assembly, it was assumed that

the buffer conditions for the individual E2 core would give optimum conditions for overall complex assembly. With a 24-mer E2 core, the assembled complex was expected to have a molecular mass of ~ 5 MDa. Theoretically, if all 24 peripheral subunit binding domain sites were occupied with either E1 (165 kDa) or E3 (99.7 kDa) the average molecular mass estimate could be ~ 5 MDa and ~ 3.3 MDa, respectively. Mixtures containing all three BCOADHC proteins in a E1:E2:E3 molar ratio of 3:1:0.1 were prepared with and without NaCl during the heat incubation step, as well as mixing the three component proteins, which had been previously incubated with NaCl individually (Fig. 9.3). At 25°C the estimated mass of the whole complex varied greatly between $6.5x10^6$ Da and $11.3x10^6$ Da, whereas DLS measurements at 55°C gave a molecular mass estimate for the whole complex centred around 5 MDa, independently of how the sample was prepared (Fig. 9.3 and Table 9.2).

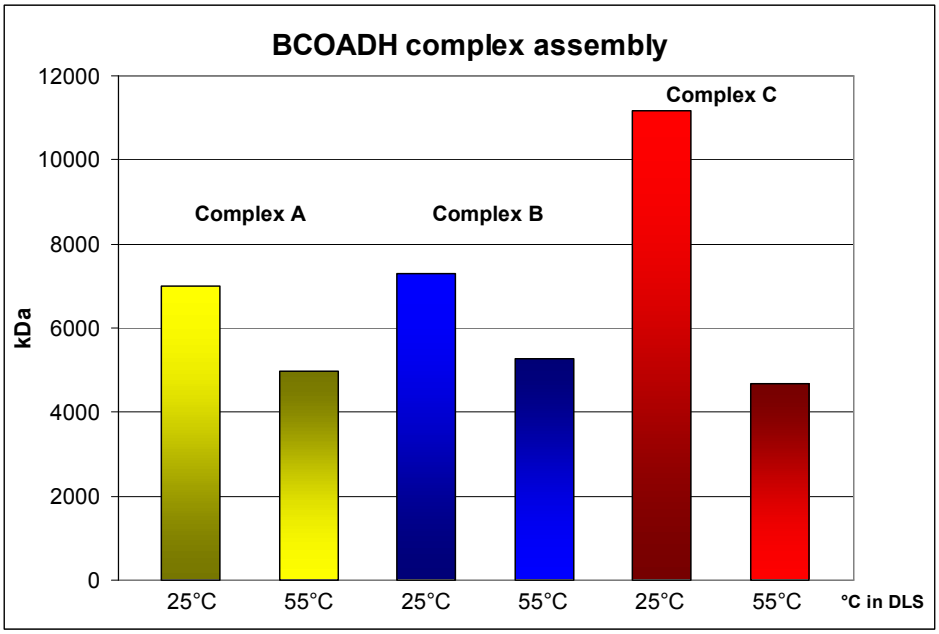


Fig. 9.3 DLS analysis of BCOADHC assembly at 25°C and 55°C. The complex was assembled in the 3:1:0.1 (E1:E2:E3) molar ratio under three different conditions and analysed in DLS at 25°C or 55°C. **Complex A** (yellow): Complex proteins were mixed and incubated at 55°C for 10 min without extra NaCl added; NaCl was supplemented prior to the DLS experiment. Estimated mass: $7x10^6$ Da (25°C) and $5x10^6$ Da (55°C). **Complex B** (blue): Complex proteins were mixed and incubated at 55°C for 10 min with 100 mM NaCl. Estimated mass: $7.3x10^6$ Da (25°C) and $5.3x10^6$ Da (55°C). **Complex C** (red): Complex proteins were incubated individually at 55°C with 100 mM NaCl and then mixed in the 3:1:0.1 ratio. Estimated mass: $11.2x10^6$ Da (25°C) and $4.7x10^6$ Da (55°C).

Table 9.2 Summary of the molecular mass estimates of whole BCOADHC assembly. **Complex A**: Proteins were mixed and incubated at 55°C for 10 min without NaCl. 100 mM NaCl was added after the mixture had cooled down. **Complex B**: Proteins were mixed and incubated at 55°C with 100 mM NaCl. **Complex C**: Proteins were first incubated at 55°C with 100 mM NaCl individually and then mixed. DLS measurements were done at 25°C and 55°C. Molecular mass estimates are in MDa ± SE.

	NaCl (100 mM) in heat incubation	DLS measurement 25°C	DLS measurement 55°C
Complex A	Absent	7.0 ± 0.5	5.0 ± 0.2
Complex B	Present	7.3 ± 0.45	5.3 ± 0.1
Complex C	Present	11.2 ± 0.3	5.5 ± 0.3

The overall results from DLS indicated that temperature is essential to drive E2 assembly and maintain whole complex assembly. Moreover, the optimum temperature for complex formation was 55°C, which matches the growth temperature of *Tp. acidophilum*.

9.2.3 Analytical ultracentrifugation of the BCOADHC

All previous attempts to characterise the *Thermoplasma* BCOADHC by analytical ultracentrifugation (AUC) were unsuccessful; the major obstacle was the presence of large amounts of individual complex components and/or protein self-oligomerization (Caroline Heath, University of Bath, personal communication). The DLS results suggested that temperature is not only a driving factor in complex assembly, but also in maintaining the overall complex with an average molecular size of ~ 5 MDa. The maximum permissible temperature for the AUC experiment on the available instrument was 40°C. Since 40°C is the critical temperature at which E2 core assembly can take place, it was decided to run AUC at 40°C. Furthermore, previous AUC analyses included 0.4 M NaCl which might have contributed to protein self-association as observed in DLS (Fig. 9.1). Therefore, the complex was prepared in the same conditions as for DLS (Fig. 9.3, Table 9.2). Complex A and B did not give sufficient data in AUC experiments. However, sedimentation velocity analysis at 40°C of complex C revealed three discrete protein peaks with s°_{20w} values of 50S (24% of total protein), 19S (8%) and 6S (68%) (Fig. 9.4). Attempts to fit these data using the

finite element solution of the Lamm equation in Sedfit (Schuck, 1998) were unsuccessful in producing M_r values for these proteins. Using the relationship that $s \propto (M_r)^{2/3}$ for globular proteins, and incorporating the values of M_r 6.1 x 10^6 and s°_{20w} value 60S for the *E. coli* PDHC (Danson *et al.*, 1979), an approximate value of $M_r = 4.7 \times 10^6$ was calculated for what is assumed to be an assembled *Tp. acidophilum* complex of 50S (Heath *et al.*, 2007).

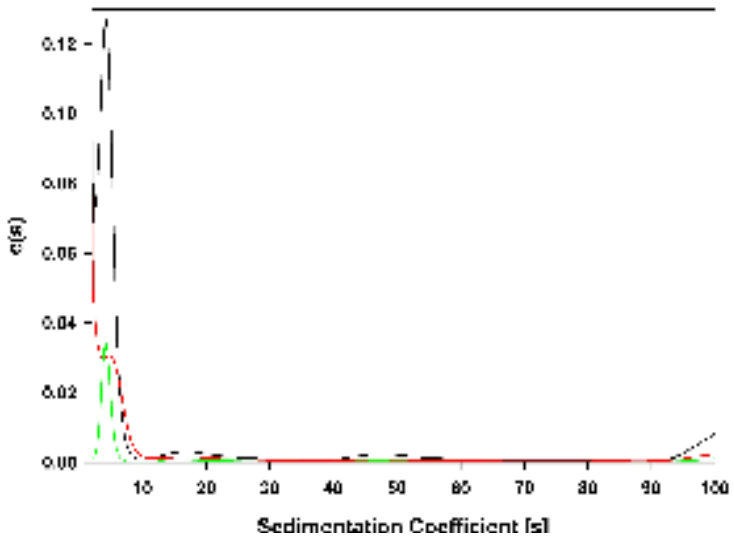


Fig. 9.4 Sedimentation velocity pattern of the whole complex (black line). Previous to AUC, proteins were incubated at 55°C with 0.1 M NaCl individually and then mixed in the molar ratio E1:E2:E3 of 3:1:0.1. The area under the peaks represents the total amount of protein.

9. 3 Discussion and conclusions

DLS and AUC were successfully employed to determine the complex assembly of the *Tp. acidophilum* BCOADHC. The DLS experiments revealed the need for high temperature to drive assembly of the E2 core and whole complex. This is an observation that has not been reported for other thermophilic OADHCs as yet. On the contrary, the recombinantly expressed PDHC from *Bacillus stearothermophilus* assembles at 4°C *in vitro* (Domingo *et al.*, 1999; Milne *et al.*, 2006) despite the organism's optimum growth temperature of 60°C. This spontaneous self-assembly along with structural features indicated that the *B. stearothermophilus* PDHC may resemble mesophilic counterparts (Witzmann and Bisswanger, 1998). In contrast, the *T. acidophilum* complex requires a critical temperature that drives the assembly of the E2 core and maintains the

whole complex assembly. In the DLS experiments with E2, which was not subjected to heat-precipitation purification, it was observed that the temperature is required for assembly and, once assembled, the E2 core remains intact even at 25°C. This suggests that the interaction between the individual E2 subunits is stable and favourable. The E2 core stability, as soon as it has assembled, is further illustrated by the observation that E2 purified from cell lysate by heat-precipitation at 65°C for 5 min and kept at 4°C before AUC analyses was also found to be present as a 24-mer core at 20°C (Caroline Heath, Ph.D. Thesis, University of Bath, 2006). Nonetheless, AUC analysis of the whole complex required a constant high temperature. This suggests that the binding of the E1 and E3 components to the E2 core may also be driven and/or maintained by temperature. This indicates that these proteins can work well at 55°C, which is commensurate with the optimum growth temperature of the *T. acidophilum*.

Several possibilities may explain the observed temperature dependence of E2 core and complex assembly. One explanation may be that the proteins are not in their proper structural conformation at lower temperatures which prevents subunit interactions. Another aspect is that a higher temperature *i.e.* input of energy, may account for the enthalpy cost for expelling water molecules (desolvation barrier) between interacting protein subunits (Levy and Onuchi, 2006). Temperature dependent oligomerization has been reported for much smaller proteins from hyperthermophlic archaea including an L-isoaspartyl-O-methyltransferase from *Sulfolobus tododaii* (Tanaka *et al.*, 2004), a recombinant amidase from *Sulfolobus solfataricus* (Politi *et al.*, 2009) and a glutamate dehydrogenase from *Pyrococcus* KOD1 (Rahman *et al.*, 1997). In these studies oligomerization is assumed to enhance protein thermostabilitly. Further investigations are required to establish how exactly temperature affects the E2 conformation and subunit-subunit interactions of the *Thermoplasma* BCOADHC. Structural information is required to characterize these interactions.

The molecular mass estimates of the complex obtained from DLS and AUC are in good agreement with each other. Both techniques reveal a complex of approximately 5 MDa. However, it is difficult to estimate the physiological ratio of E1:E2:E3 constituting the complex and experimental conditions have been

shown to affect the ratio of E1:E3 bound to the E2 core (Domingo *et al.*, 1999). The optimum ratio for *Thermoplasma* OADHC activity has been determined to be 3:1:01 for E1:E2:E3 (Heath *et al.*, 2007). However, analytical gel filtration analysis of this active complex revealed that the majority of the E1 did not bind to the complex. Instead an active complex in a ratio of 1:1:0.1 of E1:E2:E3 with a size of 5 x 10^6 Da was identified (Heath *et al.*, 2007). The size estimate from DLS and the AUC experiments are in agreement with this. The large peak at 6S (Fig. 9.4) is consistent with a M_r ~150 and is thought to be unbound E1. Expression rates of the operon containing the complex genes can affect the ratio of assembled complex proteins *in vivo*. Therefore, the ratio in which the complex components assemble *in vivo* and how this corresponds to that identified as most active ratio of E1:E2:E3 of 3:1:0.1 *in vitro* (Heath *et al.*, 2007) needs attention.

Assembly of the E2 core was detected using lipoylated E2 previously (Heath *et al.*, 2007). In this work, DLS and AUC analyses of E2 core and whole complex assembly were conducted using non-lipoylated E2 and are in agreement with the result obtained using lipoylated E2. Therefore, lipoylation does not appear to be a crucial factor in complex assembly. This is not surprising, as lipoylation is a post-translational modification. However, it would be of interest to determine if the lipoylated E2 monomers would show different dynamics of E2 core assembly. An effect on the affinity of E1 and E3 to bind to lipoylated versus non-lipoylated E2 has been reported for the PDHC from the thermophilic prokaryote *B. stearothermophilus* (Domingo *et al.*, 1999). Whether lipoylation has an effect on the ratio of E1:E2:E3 in the complex of the *Tp. acidophilum* BCOADHC is an interesting aspect, which needs to be examined further. Furthermore, the ability of LplA-CTD to lipoylate the E2 monomer, the assembled E2 core and the whole assembled complex should be analyzed and compared.

It can be concluded that the *Tp. acidophilum* BCOADHC assembles as predicted from mesophilic counterparts. The temperature dependence of complex assembly suggests that this complex was designed to work at elevated temperatures. These data suggest that the complex is stable to work under native conditions and the presence of the *bcoadhc* operon along with the genes for a

lipoylation machinery suggests a potential physiological role of the BCOADHC *in vivo*.

CHAPTER 10
Conclusions and future work

The initial chapters in the thesis describe the cloning and expression of the *Tp. acidophilum lpla* and *ctd* genes. Later, the purification and characterization of the corresponding proteins, LplA and CTD have been described. A combination of both LplA and CTD have been shown to reconstitute the *Thermoplasma* lipoylation machinery, which lipoylates the E2 component of the BCOADHC *in vitro* and possibly *in vivo*. In addition to its lipoylation machinery, the temperature dependence *of the Thermoplasma* BCOADHC assembly has also been characterized. These results constitute the first report of lipoylation in an archaeon, and indicate a potential physiological role of an archaeal OADHC. Based on the knowledge gained during the course of this work, the following are some aspects of considerations that may form the basis of future research in the area.

10.1 Protein-protein interactions

The temperature dependence of the BCOADHC assembly is a novel observation for this superfamily of protein complexes (chapter 9). The identification of the specific protein subunit interactions, which determine the temperature dependence, may be a very interesting area of research in the future. With the availability of recombinantly expressed BCOADHC complex components, a model protein system has become available to study the effect of temperature on protein-protein interactions at high temperature. Elucidation of temperature dependent protein-protein interactions may better our understanding of protein stability and lead to new ideas in protein design. Apart from crystallography approach, which may be hampered by the large degree of flexibility in the complex, one emerging technique to monitor structure and dynamics is multidimensional NMR. Recently, this technique has been successfully employed to measure the assembly of a similar sized (~ 5 MDa) protein complex (Chugh *et al.*, 2008). Also, electron spray ionization mass spectrometry (ESI-MS) has emerged as a technique to study stoichiometry and interactions of macromolecular protein assemblies like the prokaryotic ribosome of over 2 MDa (Breuker and Mc Lafferty, 2008; Rose *et al.*, 2008; Benesch and Robinson, 2006). This

technique may be valuable in analyzing the E2 core and the E2 subunit-interactions, given the similarity in their M_r.

10.2 Gene order is contrary to the protein domain arrangement

In *Tp. acidophilum*, the identified lipoylation machinery requires two gene products, LplA and CTD (chapter 5), which interact with each other (chapter 6). The LplA-CTD fusion protein, in which the LplA and CTD are arranged as N- and C-terminal, respectively, as found in several other species (Cronan *et al.*, 2005), carries out lipoylation (chapter 7). Not only are the protein domains of the *Thermoplasma* lipoylation machinery encoded by two separate genes, but the gene order is also contrary to the expected protein domain order (chapter 4). It is possible that the *Tp. acidophilum* LplA domains were once encoded by a single gene, which later split up. Splitting of genes has been previously implicated in aiding the decrease in the error rate of replication at high temperatures (Snel *et al.*, 2000). However, the debate on the significance of split genes in thermophilic organisms is still ongoing (Kummerfeld and Teichmann, 2005). What cannot be denied from the presented evidence is that the LplA and CTD arrangement appears to be finely adjusted with transcriptional regulatory elements for the *lpla* gene falling within the *ctd* gene (chapter 4). This suggests that this arrangement did not happen at random, as outlined previously in chapter 4, and may provide a potential mechanism for transcriptional regulation. In this case, one has to assume that the *Tp. acidophilum* LplA has always been encoded by two separate genes. The presence of several separate genes coding for the different domains of one protein, including proteins such as RNA polymerase subunit B and alanyl-tRNA synthetase, have also been observed in the archaeon *Nanorachaeum equitans* (Randau *et al.*, 2005). Although it has become more accepted that the genomes of bacteria and archaea undergo dynamic changes (Koonin and Wolf, 2008), the reason for the presence of such split genes has been implicated in suggesting an ancestral evolutionary stage before the modern genes, which comprise all protein domains in one polypeptide (Di Giulio, 2006). If the *Thermoplasma* LplA and CTD were indeed ancestral forms of modern lipoate protein ligases, this would not only suggest that archaeal organisms were one of the first to use lipoylation, but may also imply a physiological significance of the OADHC/lipoylation system. How this correlates

to the emergence of ferredoxin oxidoreductases, the enzymes catalysing the same chemical reaction instead of OADHCs, in the overall evolutionary context would be an interesting question to pursue in future.

Another intriguing fact is that the best alignment in BLASTP searches for LplA and CTD was found to be the *Streptococcus pneumoniae* LplA. However, *S. pneumoniae* lacks OADHC activity and its E3 component is unique in having its own N-terminal lipoyl protein domain (Håkansson and Smith, 2007). A similar case of a lipoyl-domain on the E3 component has been reported for *Mycoplasma hyponeumoniae* (Matic *et al.*, 2003), but *M. hyponeumoniae* does have OADHC activity. Together with the remarkable arrangement of the *lpla* and *ctd* genes, future work will concern the evolutionary aspect of the lipoylation machinery and the potential of gene transfers. Structure determination of the complete *Tp. acidophilum* lipoylation machinery and comparison with other lipoate protein ligases will be helpful in understanding the origin of these proteins.

10.3 Gene induction by lipoic acid

Lack of lipoylation in *Thermoplasma* has been cited as a possible reason for the absence of OADHC activity *in vivo*. Contrary to this, the Western blot analysis of *Thermoplasma* cultures grown for 37 generations in media supplemented with LA has shown *in vivo* lipoylation of E2 in an archaeon (chapter 8). As BLASTP searches only identified LplA and CTD as putative lipoylation enzymes, and their lipoylation activity has been confirmed *in vitro*, the Western blot analysis suggests that LA supplement (in the media) may be the trigger for LplA-CTD activity. In the archaeon *S. solfataricus*, the regulatory response of a ß-glycosidase revealed that gene induction occurred within one generation, whereas gene repression required 38 generations in this organism (Haseltine *et al.*, 1999). Without further probing the induction mechanisms, no conclusions can be drawn for *Tp. acidophilum* at this point. To assess details at the gene level, a time course of up-regulation of LplA and CTD transcription and/or translation by LA, as well as a potential up-regulation of the OADHC, and how these genes are co-regulated should be attempted. For the archaeon *Sulfolobus solfataricus*, a genome wide transcription map has already been constructed

(Lundgren and Bernander, 2007). Techniques like DNA microarray (Butte, 2002; Nicholson *et al.*, 2004), complemented with two-dimensional difference gel electrophoresis (Komatsu *et al.,* 2006), can give the required insight into the required physiological conditions under which LplA, CTD and OADHC components are present in the cell. Similarly, if the presence of LA induces lipoylation and potentially BCOADHC activity, the interplay of the FORs becomes important. No information is available on how these two enzyme families, both of which catalyse the same chemical reaction, could co-exist. This also needs to be investigated.

10.4 Physiological role of lipoylation and BCOADHC in *Tp. acidophilum*

In addition to the thermostability aspect, the confirmation of a functional *Thermoplasma* BCOADHC is expected to fuel the debate about the existence of these genes in the relatively small *Tp. acidophilum* genome despite apparent lack of *in vivo* activity. Different *Tp. acidophilum* strains have been found in hot springs and self-heated coal refuse piles (Yasuda *et al.*, 1995). The difference and variability, as well as nutrient availability, in these growth environments may have conditioned the different *Thermoplasma* strains to respond differently to LA supply. A comparison of different strains may give insight into the evolutionary aspects of the *Tp. acidophilum* OADHC/lipoylation machinery and why the genes for these proteins are retained despite the presence of active ferredoxin oxidoreductases.

Some of the questions raised concerning the physiological significance of the OADHC/lipoylation system cannot be answered quickly and may require the development of new models. *In vivo* reporter models to conduct gene knockout experiments can give insights into the physiological significance and regulation of genes. Although genetic manipulations have been successfully conducted in *S. solfataricus* (Jonuscheit *et al.*, 2003) it is one of the few archaeal reporter systems avaible. With the isolation and characterization of plasmid pTA1 from *Tp. acidophilum*, the design of a genetic manipulation system has great potential as pointed out by Yamashiro *et al.* (2006). The development of such a system may aid in the understanding of the physiological role of the *lpla*, *ctd, oadhc* and other genes.

References

- Al Mailem D.M, Hough D.W., Danson M.J. (2008). The 2-oxoacid dehydrogenase multienzyme complex of *Haloferax volcanii*. *Extremophiles* **12**, 89-96.
- Anglister J., Grzesiek S., Ren H., Klee C.B. and Bax A. (1993). Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. *J. Biomol. NMR* **3**, 121–126
- Baldwin R.L. (1996). How Hofmeister ion interactions affect protein stability. *Biophys. J.* **71**, 2056-2063
- Bartlett M.S. (2005). Determinants of transcription initiation by archaeal RNA polymerase. *Curr. Opin. Microbiol.* **8**, 677–684
- Bennett-Lovsey RM, Hebert AD, Sternberg MJE, Kelley LA. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: Struct. Funct. Bioinf.* **70**, 611-625
- Benesch J.L.P., Robinson C.V. (2006). Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr. Opin. Struc. Biol.* **16**, 245-251
- Borukhov S., Nudler E. (2008). RNA polymerase: the vehicle of transcription. *Trends Microbiol.* **16**, 126-134
- Breuker K., McLafferty F.W. (2008). Stepwise evolution of protein native structure with electrospray in to the gas phase, 10^{-12} to 10^{2} s. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18145-18152
- Brown J.R. (2003). Ancient horizontal gene transfer. *Nature Rev. Genet.* **4**, 121-132
- Bryant P., Kriek M., Wood R.J., Roach P.L. (2006). The activity of a thermostable lipoyl synthase from *Sulfolobus solfataricus* with a synthetic octanoyl substrate. *Anal. Biochem.* **351**, 44–49
- Bunik V.I. (2003). 2-oxoacid dehydrogenase complexes in redox regulation. *Eur. J. Biochem.* **270**, 1036-1042
- Butte A. (2002). The use and analysis of microarray data. *Nat. Rev.***1**, 951-960
- Chan W.C., Salazar C., Wegiel J. (1997). Defining the tree of life: Archaea's place in the evolutionary sphere. [Online] available at http://www.nyu.edu/projects/fitch/resources/student_papers/chan.pdf (accessed May 28th, 2006)
- Chang B., Halgamuge S., Tang S.-L. (2006). Analysis of SD sequences in completed microbial genomes: Non-SD led genes are as common as SD-led genes. *Gene* **373**, 90-99
- Chernushevich I.V., Loboda A.V., Thomson B.A. (2001). An introduction to quadrupole–time-of-flight mass spectrometry. *J. Mass. Spectrom.* **36**, 849-865
- Chugh J., Sharma S., Hosur R.V. (2008). NMR insights into a megadalton-size protein assembly. *Prot. Sci.* **17**, 1319-1325
- Cobucci-Ponzano B., Conte F., Benelli D., Londei P., Flagiello A., Monti M., Pucci P., Rossi M., Moracci M. (2006). The gene of an archaeal a-L-fucosidase is expressed by translational frameshifting. *Nucleic Acids Res.* **34**, 4258–4268
- Corpet F. (1988). Multiple sequence alignment with hierarchical clustering. *Nuceic. Acids Res.*, **16**, 10881-10890
- Cronan J.E., Zhao X., Jiang Y. (2005). Function, attachment and synthesis of lipoic acid in *Escherichia coli*. *Adv. Microb. Physiol.* **50**, 103-146
- Danson M.J., Eisenthal R., Hall S., Kessell S.R., Williams D.L. (1984). Dihydrolipoamide dehydrogenase from halophilic archaebacteria. *Biochem. J.* **218**, 811-818

- Danson M.J., Fersht A.R., Perham R.N. (1978b). Rapid intramolecular coupling of active sites in the pyruvate dehydrogenase complex of *Escherichia coli*: Mechanisms for rate enhancement in a multimeric structure. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5386-5390
- Danson M.J., Hale G., Johnson P., Perham R.N., Smith J., Spragg P. (1979). Molecular weight and symmetry of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *J. Mol. Biol.* **129**, 603-617
- Danson M.J., Hooper E.A., Perham R.N. (1978a). Intramolecular coupling of active sites in the pyruvate deyhdrogenase multienzyme complex of *Escherichia coli*. *Biochem. J.* **175**, 193-198
- Danson M.J., Morgan D.J., Jeffries A.C., Hough D.W., Dyall-Smith M.L. (2004). Multienzyme complexes in the Archaea: predictions from genome sequences. In: Halophilic microorganisms (Ventosa A., ed.), pp. 177-191. Springer Verlag
- Di Giulio M. (2006). *Nanoarchaeum equitans* is a living fossil. *J. Theoret. Biol.* **242**, 257-260
- Domingo G.J., Chauhan H.J., Lessard I.A.D., Fuller C., Perham R.N. (1999). Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*. *Eur. J. Biochem*, **266**, 1136-1146
- Dworking J., Losick R. (2001). Linking nutritional status to gene activation and development. *Genes Dev.* **15**, 1051-1054
- Forterre P., Brochier C., Philippe H. (2002). Evolution of the Archaea. *Theor. Popul. Biol.* **61**, 409-422
- Fries M., Stotta K.M., Reynolds S., Perham R.N. (2007). Distinct modes of recognition of the lipoyl domain as substrate by the E1 and E3 components of the pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* **366**, 132-139
- Fujiwara K., Okamura-Ikeda K., Motokawa Y. (1995). Assay for the protein lipoylation reaction. *Methods Enzymol.* **251**, 340-347
- Fujiwara F., Toma S., Okamura-Ikeda K., Motokawa Y., Nakagawa A., Taniguchi H. (2005). Crystal structure of lipoate-protein ligase A from *Escherichia coli*: Determination of the lipoic acid-binding site. *J. Biol. Chem*., **280**, 33645-33651
- Fujiwara F., Hosaka H., Matsuda M., Okamura-Ikeda K., Motokawa Y., Suzuki M., Nakagawa A., Taniguchi H. (2007). Crystal structure of bovine lipoyltransferase in complex with lipoyl-AMP. *J. Mol. Biol.* **371**, 222-34
- Giedroc D.P., Cornish P.V (2008). Frameshifting RNA pseudoknots: Structure and mechanism. V*irus Res.* **139**, 193-208
- Gouet P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305-8
- Gregor D., Pfeifer F. (2005). *In vivo* analyses of constitutive and regulated promoters in halophilic archaea. *Microbiol.* **151**, 25-33
- Günther S., McMillan P.J., Wallace L.J.M., Müller S. (2005). *Plasmodium falciparum* possesses organelle-specific α-keto acid dehydrogenase complexes and lipoylation pathways. *Biochem. Soc. Trans.* **33**, 977-980
- Håkansson A., Smith A.W. (2007). Enzymatic characterization of dihydrolipoamide dehydrogenase from *Streptococcus pneumoniae* harboring its own substrate. *J. Biol. Chem.* **282**, 29521-29530
- Hall M.N., Gabay J., Débarbouillé M., Schwartz M. (1982). A role for mRNA secondary structure in the control of translation initiation. *Nature* **295**, 616-618
- Haseltine C., Montalvo-Rodriguez R., Bini E., Carl A., Blum P. (1999). Coordinate transcriptional control in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* **181**, 3920–3927

- Heath C. (2006). A branched-chain 2-oxoacid dehydrogenase multienzyme complex from *Thermoplasma acidophilum*. Ph.D. Thesis, University of Bath.
- Heath C., Jeffries A.C., Hough D.W., Danson M.J. (2004). Discovery of the catalytic function of a putative 2-oxoacid dehydrogenase multienzyme complex in the thermophilic archaeon *Thermoplasma acidophilum*. *FEBS Lett.* **577**, 523-527
- Heath C., Posner M.G., Aass, H.C., Upadhyay A., Scott D.J., Hough D.W., Danson M.J. (2007). The 2-oxoacid dehydrogenase multi-enzyme complex of the archaeon *Thermoplasma acidophilum* - recombinant expression, assembly and characterization. *FEBS J.* **274**, 5406-5415
- Howard M.J., Chauhan H.J., Domingo G.J., Fuller C., Perham R.N. (2000). Protein-protein interaction revealed by NMR T2 relaxation experiments: The lipoyl domain and E1 component of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *J. Mol. Biol.* **295**, 1023-1037
- Iwabe,N., Kuma,K., Hasegawa,M., Osawa,S., Miyata,T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355-9359
- Izard T., Ævarsson A., Allen M.D., Westphal A.H., Perham R.N. de Kok A., Hol W.G.J. (1999). Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1240-1245
- Jenal U., Thurner C., Leisinger T. (1993). Transcription of the ileS operon in the archaeon *Methanobacterium thermoautotrophicum Marburg*. *J Bacteriol.* **175**, 5945–5952
- Jolley K.A., Maddocks D.G., Gyles S.L., Mullan Z., Tang S.-L., Dyall-Smith M.L., Hough D.W., Danson M.J. (2000). 2-oxoacid dehydrogenase multienzyme complexes in the halophilic Archaea? – Gene sequences and protein structural predictions. *Microbiology* **146**, 1061-1069
- Jolley K.A., Rapaport E., Hough D.W., Danson M.J., Woods W.G., Dyall-Smith M.L. (1996). Dihydrolipoamide dehydrognase form the halophilic archaeon *Haloferax volcanii*: homologous overexpression of the cloned gene. *J. Bacteriol.* **178**, 3044-3048
- Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195-202.
- Jonuscheit M., Martusewitsch E., Stedman K.M., Schleper C. (2003). A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* **48**, 1241-1252
- Jordan S.W., Cronan J.E. Jr. (2003). The *Escherichia coli* lipB gene encodes lipoyl (octanoyl)-acyl carrier protein: protein transferase. *J. Bacteriol.* **185**, 1582-1589
- Jung H-I., Cooper A., Perham R.N. (2002). Identification of key amino acid residues in the assembly of enzymes into the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*: A kinetic and thermodynamic analysis. *Biochemistry* **41**, 10446-10453
- Jung H-I., Cooper A., Perham R.N. (2003). Interactions of the peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component in the assembly of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *FEBS J.* **270**, 4488-4496
- Jürgens G. (2002). [Online] available at http://ethesis.helsinki.fi/julkaisut/maa/skemi/vk/jurgens/molecula.pdf (accessed May 28th, 2006)

- Kanai A., Sato A., Imoto J., Tomita M. (2006). Archaeal *Pyrococcus furiosus* thymidylate synthase 1 is an RNA-binding protein. *Biochem J.* **393**, 373-379
- Kang S.G., Jeong H.K., Lee E., Natarajan S. (2007). Characterization of a lipoate-protein ligase A gene of rice (*Oryza sativa L.*). *Gene* **393**, 53-61
- Karlin S., Mrázek J., Ma J., Brocchieri L. (2005). Predicted highly expressed genes in archaeal genomes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7303-7308
- Kelly S.M., Jess T.J., Price N.C. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**,119 – 139
- Kerscher L., Nowitzki S., Oesterhelt D. (1982). Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid:ferredoxin oxidoreductases. *Eur. J. Biochem.* **128**, 223-230
- Kerscher L., Oesterhelt D. (1981). The catalytic mechanism of 2-oxoacid:ferredoxin oxidoreductases from *Halobacterium halobium* - one-electron transfer at two distinct steps of the catalytic cycle. *Eur. J. Biochem.* **116**, 595-600
- Kim D.J., Kim K.H., Lee H.H., Lee S.J., Ha J.Y., Yoon H.J., Suh S.W. (2005). Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. *J. Biol. Chem.* **280**, 38081-38989
- Komatsu S., Zang X., Tanaka N. (2006). Comparison of two proteomics techniques used to identify proteins regulated by gibberellin in rice. *J. Proteome Res*. **5**, 270–276
- Koonin E. V., Wolf Y.I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688-6719
- Kosa P.F., Ghosh G., DeDecker B.S., Sigler P.B. (1997). The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor(II)B core/TATA box. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6042-6047
- Kozak M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37
- Kresge N., Simoni R.D., Hill R.L. (2006). Lester J. Reed and the α-keto acid dehydrogenase complexes. *J. Biol. Chem.* **281**, e8-e9
- Kummerfeld S.K., Teichmann S.A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **21**, 25-30
- Lee J.-I., Kang J., Stipanuk M.H. (2006). Differential regulation of glutamate–cysteine ligase subunit expression and increased holoenzyme formation in response to cysteine deprivation. *Biochem. J.* **393**, 181-190
- Levy Y., Onuchi J.N. (2006). Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389-415
- Littlefield O., Korkhin Y., Sigler P.B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13668-13673
- Lundgren M., Bernander R. (2007). Genome-wide transcription map of an archaeal cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2939–2944
- Ma J., Campbell A., Karlin S. (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733-5745
- Madhusudhan K.T., Luo J., Sokatch J.R. (1999). *In vitro* transcriptional studies of the bkd operon of *Pseudomonas putida*: L-branched-chain amino acids and D-Leucine are the inducers. *J. Bacteriol.* **181**, 2889-2894
- Matica J.N., Wiltona J.L., Towersa R.J., Scarmana A.L., Minionc F.C., Walkerb M.J., Djordjevica S.P. (2003). The pyruvate dehydrogenase complex of

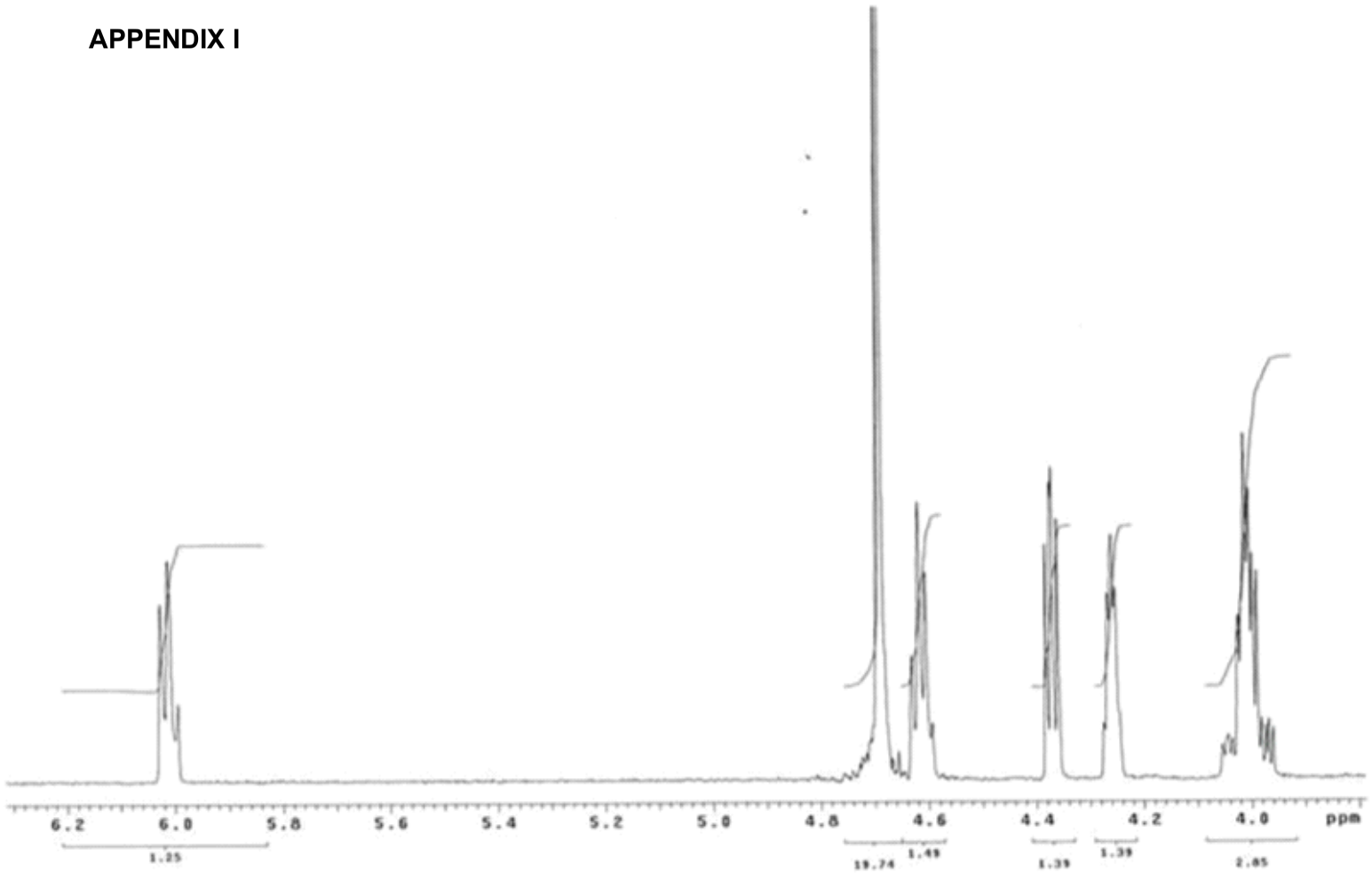
Mycoplasma hyopneumoniae contains a novel lipoyl domain arrangement. *Gene* **319**, 99-106

- McManus E., Luisi B.F., Perham R.N. (2006). Structure of a putative lipoate protein ligase from *Thermoplasma acidophilum* and the mechanism of target selection for post-translational modification. *J. Mol. Biol.* **356**, 625-637
- McMurry J. (1999). Organic Chemistry, 5th edition. Pacific Grove,CA 93950, USA: Brooks/Cole.
- Miles J.S, Guest J.R., Radford S.E., Perham R.N. (1988). Investigation of the mechanism of active site coupling in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* by protein engineering. *J. Mol. Biol.* **202**, 97-106
- Milne J.L.S., Shi D., Peter B.Rosenthal P.B., Sunshine J.S., Domingo G.J., Wu X., Bernard R.Brooks B.R., Perham R.N., Richard Henderson R., Subramaniam S. (2002). Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *EMBO J.* **21**, 5587-5598
- Milne J.L.S., Wu X., Borgnia M.J., Lengyel J.S., Brooks B.R., She D., Perham R.N., Subramaniam S. (2006). Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. *J. Biol. Chem.* **281**, 4364-4370
- Morris T.W., Reed K.E., Cronan J.E. (1994). Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. *J. Biol. Chem.* **269**, 16091-16100
- Nguyen B., Farial A. Tanious, W. David Wilson (2007). Biosensor-surface plasmon resonance: Quantitative analysis of small molecule–nucleic acid interactions. *Methods* **42**, 150-161
- Nicholson T.L., Chiu K., Stephens R.S. (2004). *Chlamydia trachomatis* lacks an adaptive response to changes in carbon source availability. *Infect. Immun.* **72**, 4286-4289
- Pace N.R. (2006). Time for a change. *Nature* **441**, 289
- Park Y.-J., Yoo C-B., Choi S.Y., Lee H.-B. (2006). Purifications and Characterizations of a ferredoxin and its related 2-oxoacid:ferredoxin oxidoreductase from the hyperthermophilic archaeon *Sulfolobus solfataricus P1*. *J. Biochem. Mol. Biol.* **39**, 46-54
- Perham R.N. (1975). Self-assembly of biological macromolecules. *Philos. Trans. R. Soc. London [Biol]* **272**, 123-136
- Perham R.N. (1991). Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: A paradigm in the design of a multifunctional protein. *Biochemistry* **30**, 8501-8512
- Perham R.N. (2000). Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Ann. Rev. Biochem.* **69**, 961-1004
- Perham R.N., Packman L.C. (1989). 2-Oxo acid dehydrogenase multienzyme-complexes: domains, dynamics, and design. *Ann. NY Acad. Sci.* **573**,1-20
- Perham R.N., Thomas J.O. (1971). The subunit molecular weights of the α-ketoacid dehydrogenase multienzyme complexes from *E. coli*. *FEBS Lett*. **15**, 8-12
- Politi L., Chiancone E. Giangiacomo L., Cervoni L., D'abusco A.S., Scorsino S., Scandurra R. (2009). pH-, temperature- and ion-dependent oligomerization of *Sulfolobus solfataricus* recombinant amidase: a study with site-specific mutants. *Archaea* **2**, 221-231
- Quinn J., Diamond A.G., Masters A.K., Brookfield D.E., Wallis N.G., Yeaman S.J. (1993). Expression and lipoylation in *Escherichia coli* of the inner lipoyl domain of the E2 component of the human pyruvate dehydrogenase complex. *Biochem. J.* **289**, 81-85

- Qureshi S.A., Jackson S.P. (1998). Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* **1**, 389-400
- Rahman R.N.Z.A., Fujiwara S., Takagi M., Kanaja S., Imanaka T. (1997). Effect of heat treatment on proper oligomeric structure formation of thermostable glutamate dehydrogenase from a hyperthermophilic archaeon. *Biochem. Biophys. Res. Comm.* **241**, 646-652
- Randau L., Münch R., Hohn M.J., Jahn D., Söll D. (2005). *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature*, **433**, 537-541
- Reed L.J., Leach F.R., Koike M. (1958). Studies on a lipoic acid-activating system. *J. Biol. Chem.* **232**, 123-142
- Reed, L.J., DeBusk, B.G., Gunsalus, I.C., Hornberger, C. S., Jr. (1951). Crystalline α-lipoic acid: a catalytic agent associated with pyruvate dehydrogenase. *Science* **114**, 93–94
- Reeder J.,Giegrich R. (2004). Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics* **5,** 104
- Rose R. J., Verger D., Daviter T., Remaut H., Paci E., Waksman G., Ashcroft A.E., Radford S.E. (2008). Unravelling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12873-12878
- Schreiner M.E. Fiur D., Holátko J., Pátek M., Eikmanns B.J. (2005). E1 enzyme of the pyruvate dehydrogenase complex in *Corynebacterium glutamicum*: molecular analysis of the gene and phylogenetic aspects. *J. Bacteriol.* **187**, 6005–6018
- Schuck P. (1998). Sedimentation analysis of non-interacting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503–1512.
- Schuck P. (2000). Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619
- Self W.T., Tsai L., Stadtman T.C. (2000). Synthesis and characterization of selenotrislide-derivatives of lipoic acid and lipoamide. *Proc. Natl. Acad. Sci. U.S.A.* **97** (23), 12481-12486
- Smith A.W., Roche H., Trombe M., Briles D.E., Håkansson A. (2002). Characterization of the dihydrolipoamide dehydrogenase from *Streptococcus pneumoniae* and its role in pneumococcal infection. *Mol. Microbiol.* **44**, 431-448
- Smith L.D., Bungard S.J., Danson M.J., Hough D.W. (1987). Dihydrolipoamide dehydrogenase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum. Biochem. Soc. Trans.* **15**, 1097
- Snel B., Bork P., Huynen M. (2000). Genome evolution: gene fusion versus gene fission. *Trends Genet.* **16**,
- Stanley C.J., Packman L.C., Danson M.J., Henderson C.E., Perham R.N. (1981). Intramolecular coupling of active sites in the pyruvate dehydrogenase multienyzme complexes from bacterial and mammalian sources. *Biochem. J.* **195**, 715-721
- Stepp L.R., Bleile D.M., McRorie D.K., Pettit F.H., Reed L.J. (1981). Use of trypsin and lipoamidase to study the role of lipoic acid moieties in the pyruvate and α-ketoglutarate dehydrogenase complexes of *Escherichia coli*. *Biochemistry* **20**, 4555-4560
- Sun N., Beck F., Knispel R.W., Siedler F., Scheffer B., Nickell S., Baumeister W., Nagy I. (2007). Proteomics analysis of *Thermoplasma acidophilum* with a focus on protein complexes. *Mol. Cell. Proteomics* **6**, 492-502

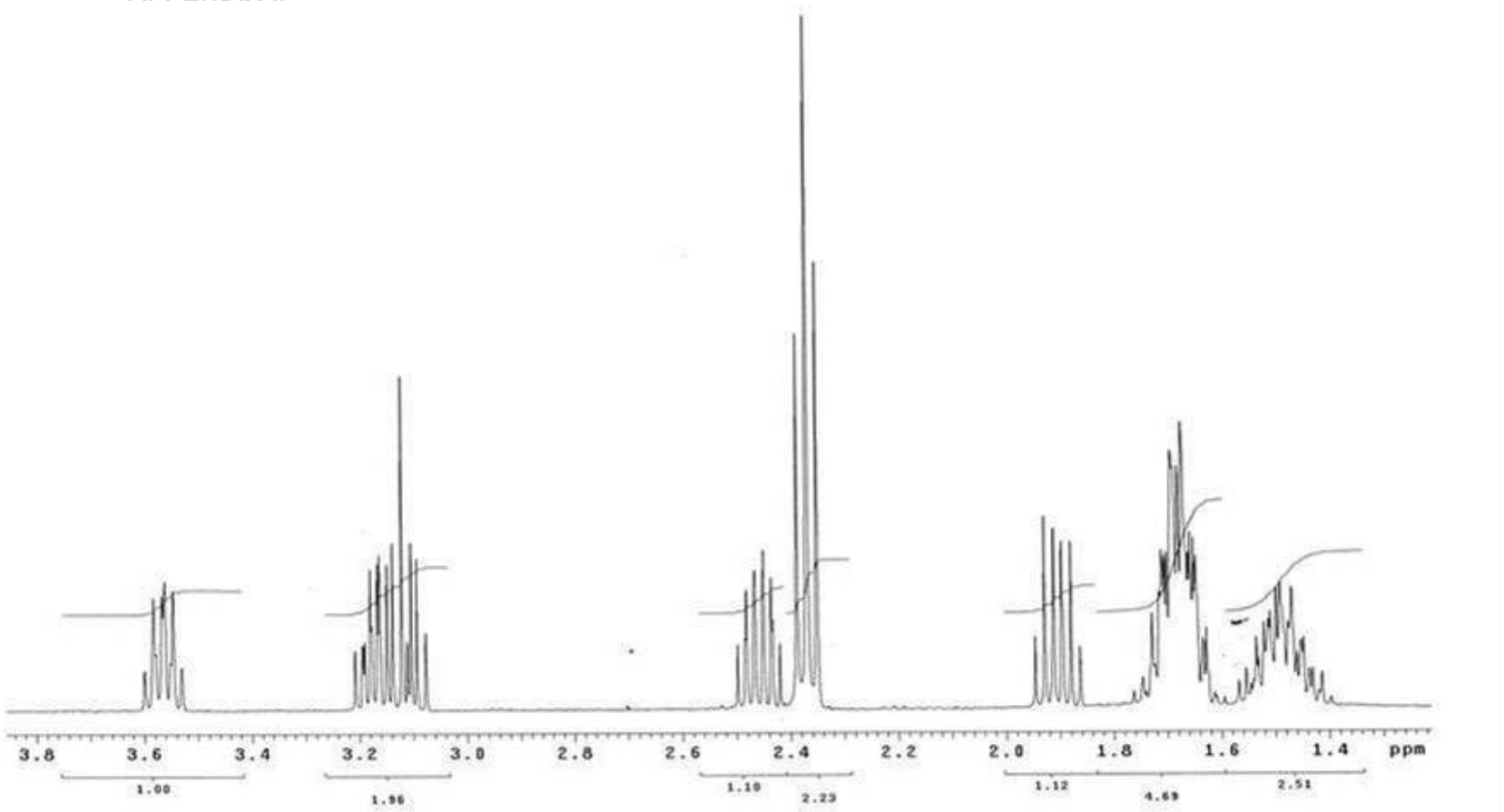
- Tanaka Y., Tsumoto K., Yasutake Y., Umetsu M., Yao M., Fukada H., Tanaka I., Kumagai I. (2004). How Oligomerization contributes to the thermostability of an archaeon protein. *J. Biol. Chem.* **279**, 32957–32967
- Tang X., Ezaki S., Fujiwara S., Takagi M., Atomi H. and Imanaka T. (1999). The tryptophan biosynthesis gene cluster trpCDEGFBA from *Pyrococcus kodakaraensis* KOD1 is regulated at the transcriptional level and expressed as a single mRNA. *Mol. Gen. Genet.* **262**, 815-821
- Thevelein J.M., Bonini B.M., Castermans D., Haesendonckx S., Kriel J., Louwet W., Thayumanavan P., Popova Y., Rubio-Texeira M., Schepers W., Vandormael P., Van Zeebroeck G., Verhaert P., Versele M., Voordeckers K. (2008). Novel mechanisms in nutrient activation of the yeast protein kinase A pathway. *Acta. Microbiol. Immunol. Hung.* **55**,75-89.
- Thomas A.S., Elcock A.H. (2006). Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations: differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson-Boltzmann theory. *J. Am. Chem. Soc.* **128**, 7796-7806
- Torarinsson E., Klenk H.-P., Garrett R.A. (2005). Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* **7**, 47-54
- Touchstone J.C. (1978). Practice of thin layer chromatography. New York; Chichester; Wiley
- van Ooyen J., Soppa J. (2007). Three 2-oxoacid dehydrogenase operons in *Haloferax volcanii*: expression, deletion mutants and evolution. *Microbiology* **153**, 3303–3313
- van Ooyne J., Soppa J. (2007). Three 2-oxoacid dehydrogenase operons in *Haloferax volcanii*: expression, deletion mutants and evolution. *Microbiology* **153**, 3303–3313
- Varki A., Cummings R.D., Esko J.D., Freeze H.H., Stanley P., Bertozzi C.R., Hart G.W., Etzler M.E. eds. (2009). Essentials of glycobiology. 2nd ed. Califronia: Cold Spring Harbor Press.
- Vieille C., Zeikus G.J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1-42
- Wanner, C., Soppa, J. (2002). Functional role for a 2-oxo acid dehydrogenase in the halophilic archaeon *Haloferax volcanii*. *J. Bacteriol*. **184**, 3114–3121
- Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* **65**, 1395-1404
- Witzmann S., Bisswanger H. (1998). The pyruvate dehydrogenase complex from thermophilic organisms: thermal stability and re-association from the enzyme components. *Biochim. Biophys. Acta* **1385**, 341-352
- Woese,C.R., Kandler,O., Wheelis,M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 4576-4579
- Yamashiro K., Yokobori S., Oshima T., Yamagishi A. (2006). Structural analysis of the plasmid pTA1 isolated from the thermoacidophilic archaeon *Thermoplasma acidophilum*. *Extremophiles* **10**, 327–335
- Yasuda M., Oyaizu H., Yamagishi A., Oshima T. (1995). Morphological variation of new *Thermoplasma acidophilum* isolates from Japanese hot springs. *Appl. Env. Microbiol.* **61**, 3482-3485
- Yeaman S.J. (1989). The 2-oxo acid dehydrogenase complexes: recent advances. *Biochem J.* **257**, 625-632
- Zhao X., Miller J.R., Jiang Y., Marletta M.A., Cronan J.E. (2003). Assembly of the covalent linkage between lipoic acid and its cognate enzymes. *Chem. Biol.* **10**, 1293-1302

APPENDIX I



^{1}H NMR spectrum of AMP. Solvent D_2O.

APPENDIX II



^{1}H NMR spectrum of lipoic acid. Solvent $CDCl_3$.

APPENDIX III